医药新科技丛书

丛书主编 朱作言

基因分析和生物芯片技术

GENE ANALYSIS AND BIOCHIP TECHNIQUE

主编 丁金凤 杨渝珍 张先恩

湖北科学技术出版社

GENE ANALYSIS AND BIOCHIP TECHNIQUE

策　　划　冯友仁
责任编辑　冯友仁
美术编辑　戴　旻

ISBN 7-5352-2933-6
9 787535 229335 >

ISBN 7-5352-2933-6
R · 659　定价：58.00元

Q78
D513

医药新科技丛书

丛书主编 朱作言

基因分析和生物芯片技术

GENE ANALYSIS AND BIOCHIP TECHNIQUE

主编 丁金凤 杨渝珍 张先恩

A1110648

華僑大學圖書館藏書

湖北科学技术出版社

图书在版编目(CIP)数据

基因分析和生物芯片技术/丁金凤等主编.—武汉:湖北科学技术出版社,2004.1
(医药新科技丛书/朱作言主编)
ISBN 7-5352-2933-6

Ⅰ.基… Ⅱ.丁… Ⅲ.①基因—分析②生物—芯片 Ⅳ.Q78

中国版本图书馆 CIP 数据核字(2003)第 074709 号

医药新科技丛书
基因分析和生物芯片技术
© 丁金凤 杨渝珍 张先恩 主编
策 划:冯友仁 封面设计:戴 旻
责任编辑:冯友仁

出版发行:湖北科学技术出版社 电话:87679468
地 址:武汉市雄楚大街 268 号湖北出版文化城 B 座 12-14 层 邮编:430070

印 刷:湖北新华印务有限公司 邮编:430034
督 印:刘春尧

787 毫米×1092 毫米 16 开 26 印张 530 千字
2004 年 1 月第 1 版 2004 年 1 月第 1 次印刷

ISBN 7-5352-2933-6/R·659 定价:58.00 元

本书如有印装质量问题 可找承印厂更换

编委名单及工作单位

中国医学科学院心血管病研究所	丁金凤　李　健　刘冬青 孟宪敏　杨菊先　陈兰英 赵　勇　曹慧青　潘德思
华中科技大学同济医学院	杨渝珍　杨庆恩　陈　慧 徐顺清　黄　巍　朱　科 孟宪芳　赵云斌　黄代新 廖　芳　王小利　王　震 韩　玲
华中科技大学同济医院	林菊生
华中科技大学协和医院	李一荣　张德太　吴健民
华中科技大学生命科学院	骆清明
军事医学科学院生物工程研究所	杨　晓
中国科学院武汉分院病毒所	张先恩　文继开　刘　强 汪世华　陆红兵
华中农业大学	吴良志
美国耶鲁大学留美学者	马先勇
UTMD Anderson Cancer Center	魏英杰

主　编：丁金凤　杨渝珍　张先恩

副主编：陈　慧　廖　芳

编　者：（以姓氏笔画为序）

丁金凤　马先勇　王小利

王　震　文继开　李一荣

李　健　刘冬青　刘　强

朱　科　杨庆恩　杨　晓

杨渝珍　杨菊先　张先恩

张德太　陈兰英　陈　慧

孟宪芳　孟宪敏　吴良志

吴健民　陆红兵　汪世华

林菊生　赵云斌　赵　勇

骆清明　徐顺清　黄代新

黄　巍　曹慧青　韩　玲

廖　芳　潘德思　魏英杰

内容提要

本书共分为两大部分。第一部分主要介绍基因分析技术，但并不包括常规的DNA(或RNA)常规分离、纯化、基因的克隆，表达、PCR等技术；而是重点地介绍一些对未知基因的结构和功能分析时所需要的关键技术，包括cDNA文库构建、差减杂交法和差减文件的构建、mRNA差异显示技术、荧光原位杂交(FISH)技术、辐射性杂交、计算机克隆及染色体定位、核酸及蛋白序列的计算机初步分析、酵母双杂交技术、SELEX技术、基因打靶技术、转基因动物技术、RNA干扰技术，还概要地介绍了常用数据库及分子生物学软件。

作为基因分析的重要技术之，特将生物芯片技术列在第二部分进行专门介绍，共分五章，包括芯片概介、生物芯片的制备和检测、微阵列探测中的共聚焦扫描技术、生物芯片在医学中的应用(涵盖了生物芯片在基因组功能、单核苷酸多态性检测、遗传性疾病诊断、肿瘤研究、肝癌及病毒性肝炎诊断、微生物变异及耐药性检测、药学、法医学中的应用诸多领域)及蛋白质芯片。

本书编写目标定位于医学专业参考书，编写者来自生物学、医学、光电子学、化学、计算机学、生物信息学等不同的专业领域，从不同的视野来阐述一个共同的主题。在编写风格上力求深入浅出、强调图文并茂，以求一目了然，为医药学院校的研究生、教师、科研人员、临床医生等提供帮助。

序

自 DNA 双螺旋结构被发现 50 年以来，生命科学和生物技术得到了前所未有的发展，并推动相关学科向分子水平不断深入。生命科学的重大发现，以及与信息技术、生物技术、纳米技术的结合，赋予 21 世纪医药学研究旺盛的生命力，引导着理论和技术上的重大突破。由中外医学相关领域的专家和学者共同编写的《医药新科技丛书》，重点介绍了纳米技术、新基因分析技术、生物芯片技术和基因治疗技术。充分反映了前沿科学技术对医学的影响，必将为学科交叉和高新技术应用起推动作用。

纳米科学技术是 10 余年来逐渐发展起来的一项高新科学技术。著名的诺贝尔奖获得者理查德费曼早在 20 世纪 60 年代就预言，如果人类能够在原子和分子的尺度上来加工材料、制备装置，就能得到大量的异乎寻常的新发现。纳米科学技术是在纳米尺度(1～100 纳米，即 10^{-9}～10^{-7}米)研究物质新颖的物理、化学和生物特性并制造有特定功能产品的科学技术，在生物工程和医学上有光明的应用前景。科研人员已经成功地利用纳米微粒进行了细胞分离；用金纳米粒子进行病变靶向治疗，以减少副作用；用纳米微粒作为载体的病毒转导颗粒已经取得了突破性进展并用于动物实验，有望在疾病的基因治疗中发挥作用；科学家们还设想利用纳米技术制造出分子机器人，对人体的精细部位实施特殊诊断或治疗，例如修复损伤的血管壁等。希望《医药新科技丛书》中这部分内容能为相关研究领域的有志者提供第一手资料并促进纳米技术与医学的整合。

随着人类基因组精确图的绘制，越来越多的动植物、微生物基因组测序也相继完成。然而，如何去面对如此众多的基因序列，并逐一诠释他们的功能，成为全世界生命科学工作者共同面对的课题。寻找及发现新的疾病相关基因并确定它们的功能，无疑是更大的挑战；而其中隐藏了无限的商机，药物新靶标将是生物技术公司和药物开发商竞相追逐的目标。《医药新科技丛书》的编者注意到了这个问题的重要性，不吝笔墨对基因及其分析技术作了详尽描述。基因的“敲除”、“敲入”和“敲低”都是实时在体研究基因功能的关键技术。

当然，书中所提及的生物芯片更是颇有时代象征的高新技术。美国《财富》杂志将微电子芯片和生物芯片列为 20 世纪科技史上有深远影响的两件事。生物芯片技术借用计算机芯片制作理念，以高通量、高效率、自动化、微型化为特征，提供了一个从分子、基因水平向基因组、蛋白质、细胞乃至组织器官组水平回归的研究平台，有利于在整体水平上解释和探讨生命现象，具有广泛应用前景。国内已有多所大学和科研单位启动了这方面的研究，也取得了一定的进展，《医药新科技丛书》对此作了及时的反映。内容深入浅出，其主要阅览群体定位在活跃于医疗卫生领域的医生、教师、学生，兼具普及和提高的双重作用。

《医药新科技丛书》的另外一册介绍了“癌症的基因治疗”。基因治疗是一种全新的理念，折射着现代分子生物学理论和技术在医学领域中所掀起的冲击波。基因治疗的基本概念来源于人类对自身遗传本质的了解。基因作为遗传的基本单位，不仅决定着我们的相貌、高矮和气质特征，而且其细小的异常将不可避免地导致各种疾病。基因疗法的原理是用一个正常基因来置换有缺陷的基因，或者补偿缺陷基因的致病因素。自1990年第1例腺苷脱氨酶基因治疗以后的8年中，世界上已有373个临床方案被实施，累计3 134人接受了转基因试验，而且基因治疗的概念及策略已扩展到单基因遗传病的范围之外。当然，目前基因治疗的技术还有待成熟，有关安全、伦理、社会的问题还有待人们更理性地思考，希望读者和作者互动，共同促进基因治疗沿着健康的方向走向成熟。

21世纪的生命科学和技术将以难以预测的方向和难以估计的速度发展，其结果将极大地改变人类的生活方式和提高生活质量。我们面临着一个崭新的世界，《医药新科技丛书》似乎是一条小径，可以引导读者去开拓、耕耘和拥抱未来的光明。

中国科学院院士、国家自然科学基金委员会副主任

2003年5月于北京

前　言

20世纪末和21世纪初，以人类基因组计划(HGP)为标志的先进科学技术发展日新月异，PCR技术、测序技术、纳米技术、生物芯片技术、基因打靶和基因治疗技术等一个接一个地出现，生命科学正以前所未有的速度向前发展，并推动其分支学科向分子水平深入。作为21世纪的医药学研究领域，得益于基因组和后基因组时代的福荫，在疾病的诊断、治疗、预防及发病机制等各方面正都已发生、正在发生和将要发生理论和技术方面的重大突破。

我们都是从事生物化学和分子生物学领域研究和教学的，很希望能将这种进步反映和应用在我们的工作中。因此本书的编写目标定位于医学专业参考书，为医药学院校的研究生、教师、科研人员、临床医生等提供帮助。本书共分为两大部分。第一部分主要介绍基因分析技术，但是并不从基础的DNA或RNA的分离、鉴定、纯化、外源基因的克隆、表达和鉴定、PCR技术等方面入手，而是选择性地介绍一些新的基因分析技术，特别是对未知基因的结构和功能分析时需要采用的关键技术，例如从cDNA文库的构建开始，在基因组、cDNA、mRNA差减杂交或差异显示的基础上，对差异显示的片段进行计算机克隆、全长分析、辐射性杂交及染色体定位；对基因打靶技术、转基因技术、RNA干扰技术、SELEX技术也都有介绍；同时还详细介绍了基因功能研究的酵母双杂交技术。参与这部分编写的大部分人员是对这些技术的亲自操作者，因而包含了他们的经验和体会。考虑到不同层次、不同实验室的技术和设备的差异，我们特意将生物芯片技术单独组合成为第二部分，进行详细地介绍，从基本概念、定义到医学领域的应用；从基片选择、化学修饰、生物芯片种类和制作、检测流程、杂交方法，到杂交结果处理和分析，生物信息提取的原则和方法等，尽量给予深入浅出的全面的介绍。从实用的角度出发并为本书的读者群体考虑，本书并不想过多地在芯片化学修饰的分子机制、信息处理软件设计和分析原理等方面过多地染笔，只是突出生物芯片作为一种高通量、超微量的分子杂交技术和以往的经典核酸或蛋白质检测技术相比优越在何处？特点是什么？以及由此带来的技术难点和仪器设备如何与之配套等一系列问题。在弄清这些问题的基础上，本书强调了如何利用现有的，由各生物技术公司提供的生物芯片进行检测；如有条件的话，读者还可以结合自己科研的需要，制备一些特色芯片，此时也可以从本书中获得一些启发和要领。

现代科学技术是生物、化学、物理、计算机、生物信息、工程、微电子等诸多领域的综合体，学科之间的交叉、融合、渗透与整合不断地发生、发展，推动着技术的进步和新技术的诞生。因此，这本书的编写者来自医学、生物学、光电子学、化学、计算机学等不同的专业领域，从不同的视野来阐述一个共同的主题。此外，本书还得到了目前还在海外学习或工作的学子的支持和响应，及时地寄来书稿，以尽

一份拳拳之心。衷心感谢所有参与本书编写的教授、专家、博士生，特别是目前还在国外工作或学习的学者，感谢他们在繁忙工作中还挤出时间为本书撰写书稿。

本书在编写风格上力求深入浅出、强调图文并茂，为了更好地说明问题，还采用了彩图。特别在计算机克隆和生物分析软件的介绍中，均配置和下载了电脑屏幕显示，以求一目了然。

由于时间仓促，水平有限，不能满足诸位读者的需求，敬请原谅，也欢迎批评和指正。

丁金凤　杨渝珍　张先恩

2003 年 11 月于武汉

目　录

第一部分　基因分析技术

第一章　cDNA 文库的构建 …… (3)
第一节　cDNA 文库构建的基本原理 …… (3)
一、cDNA 的合成 …… (4)
二、ZAP 表达载体的结构 …… (4)
第二节　cDNA 文库构建的具体操作步骤 …… (6)
一、cDNA 第一链的合成 …… (7)
二、cDNA 第二链的合成 …… (7)
三、补平 cDNA 链末端 …… (8)
四、*EcoR* I 接头的连接 …… (9)
五、*EcoR* I 末端的磷酸化 …… (9)
六、*Xho* I 酶切 …… (9)
七、过柱分离 cDNA …… (9)
八、cDNA 与 ZAP 表达载体的连接 …… (12)
九、准备宿主菌 …… (13)
十、应用 Gigapack III Gold 包装蛋白包装含有 cDNA 的 ZAP 重组载体 …… (13)
十一、铺盘和库容量鉴定 …… (13)
十二、原始库的扩增 …… (13)

第二章　差减杂交法和差减文件的构建 …… (15)
第一节　基因组差减杂交法 …… (15)
第二节　mRNA 或 cDNA 差减杂交法 …… (16)
一、羟基磷灰石(HAP)柱层析法 …… (17)
二、生物素－链亲和素的亲和分离 …… (17)
三、基于 PCR 反应的差减杂交法或抑制性差减杂交(SSH) …… (22)
四、基因或表达序列文库(cDNA or EST libraries)杂交 …… (26)

第三章　mRNA 差异显示技术 …… (28)
第一节　mRNA DD－PCR 的基本原理 …… (28)
第二节　mRNA DD－PCR 技术的具体操作步骤 …… (29)
一、组织样品总 RNA 的提取 …… (29)
二、去除 RNA 样品中的微量 DNA …… (30)

三、引物序列 …… (31)
四、反转录归类反应体系的建立 …… (31)
五、PCR 选择性扩增 …… (32)
六、mRNA DD－PCR 产物的变性聚丙烯酰胺凝胶电泳 …… (33)
七、差异表达条带的回收 …… (33)
八、回收条带的再次扩增 …… (34)
九、二次扩增产物的克隆、测序和 GenBank 查询 …… (36)
十、Northern Blot 和 Dot Blot 杂交 …… (36)

第四章　荧光原位杂交(FISH)技术 …… (40)
第一节　FISH 的原理和技术特点 …… (40)
一、FISH 技术的原理和优点 …… (40)
二、FISH 技术的基本操作过程 …… (41)
第二节　FISH 技术的实验操作步骤 …… (43)
一、染色体标本的制备 …… (43)
二、染色体显带 …… (44)
三、Alu－PCR 法制备探针 …… (45)
四、切口平移法标记探针 …… (45)
五、FISH 检测 …… (46)
第三节　FISH 技术的发展与应用 …… (48)
一、FISH 技术的发展 …… (48)
二、FISH 技术的主要应用 …… (49)

第五章　辐射性杂交 …… (51)
第一节　辐射性杂交的基本原理 …… (51)
一、构建 RH 图谱的方法 …… (52)
二、RH 嵌板 …… (53)
第二节　RH 图谱的应用 …… (54)
一、EST 定位 …… (54)
二、寻找新基因 …… (55)
三、基因克隆 …… (55)
四、测定距离 …… (56)
五、RH 相关网站及 RH 数据库 …… (56)
六、RH 的优缺点 …… (57)

第六章　计算机克隆及染色体定位 …… (58)
第一节　计算机克隆 …… (58)
一、基本原理 …… (58)
二、现状与发展 …… (58)

三、操作步骤 …… (59)
四、一点经验 …… (59)
五、简单实例 …… (60)
第二节　染色体定位 …… (61)
第三节　掌握和熟悉 NCBI 数据库和分析 …… (64)

第七章　核酸及蛋白序列的计算机初步分析 …… (66)
第一节　概述 …… (66)
第二节　核酸序列分析 …… (69)
一、数据库同源性或相似性搜索 …… (69)
二、读码框分析 …… (69)
三、寻找启动子区及其核酸序列中的功能位点 …… (70)
四、基因的染色体定位 …… (70)
第三节　蛋白序列分析 …… (70)
一、同源性或相似性分析 …… (70)
二、氨基酸序列的一般特性分析 …… (71)
三、蛋白基序和结构域分析 …… (71)
四、特殊结构或结构特征分析 …… (71)
五、直系同源物(Ortholog)和/或旁系同源物(Paralog)分析 …… (72)
六、二级结构的预测 …… (72)
第四节　多重序列排列和系统发育分析 …… (73)
一、多重序列排列 …… (73)
二、分子进化分析 …… (73)
三、举例 …… (75)
第五节　刊有生物信息学的期刊 …… (79)

第八章　酵母双杂交技术 …… (80)
第一节　酵母双杂交原理 …… (80)
一、原理 …… (80)
二、酵母双杂交的基本流程 …… (81)
第二节　酵母双杂交系统中常用的菌株与表达载体 …… (82)
一、酵母菌的一般特点 …… (82)
二、常用的菌株 …… (82)
三、表达载体的构建 …… (83)
四、表达载体与菌株表型之间的关系 …… (85)
五、对照载体的构建 …… (85)
六、酵母培养基的制备 …… (86)
七、菌株冻存与复苏 …… (88)
八、酵母菌的培养 …… (89)

第三节　诱饵质粒的鉴定 …… (89)
一、制备感受态酵母菌 …… (89)
二、诱饵质粒转化入感受态酵母细胞 …… (90)
三、诱饵蛋白的毒性检测 …… (90)
四、诱饵蛋白自激活检测 …… (91)
五、交配效率的检测与计算 …… (92)
第四节　正式筛选预转化 cDNA 文库 …… (93)
一、筛选前的准备 …… (93)
二、正式筛选预转化的 cDNA 文库 …… (94)
三、文库滴定(Library Titering) …… (96)
四、计算酵母交配的效率和筛选的克隆数 …… (96)
第五节　对初步获取的阳性克隆的分析与验证 …… (97)
一、克隆表型验证——大规模筛选 …… (97)
二、β－半乳糖苷酶活性分析 …… (97)
三、酵母质粒 DNA 的提取 …… (99)
四、酵母质粒转入大肠杆菌 DH5α …… (100)
五、克隆归类 …… (101)
六、在酵母体内再验证蛋白的相互作用 …… (102)
七、在酵母体外分析和验证蛋白的相互作用 …… (103)
八、测序 …… (103)
第六节　酵母双杂交技术的应用 …… (103)
一、酵母双杂交技术的应用 …… (103)
二、酵母双杂交技术的缺陷 …… (104)
三、使用酵母双杂交技术应注意的问题 …… (104)

第九章　SELEX 技术 …… (105)
第一节　SELEX 技术简介 …… (105)
第二节　SELEX 技术的基本原理 …… (106)
一、SELEX 技术的基本路线 …… (106)
二、Aptamer 的修饰与优化 …… (108)
三、影响 SELEX 技术的因素 …… (111)
第三节　核酸配基结合的特异性评价 …… (116)
一、配基结合特异性 …… (116)
二、配基序列及空间结构分析 …… (117)
第四节　SELEX 技术的改进与发展 …… (118)
一、导向 SELEX …… (118)
二、复合靶分子 SELEX …… (118)
三、基因组 SELEX …… (119)
第五节　适配分子的特性 …… (119)

第六节　SELEX 技术和适配分子在各方面的应用 …… (120)
一、SELEX 技术在各方面的应用 …… (120)
二、Aptamer 在各方面的应用 …… (122)

第十章　基因打靶技术 …… (135)
第一节　基因打靶技术的概念及发展历史 …… (135)
一、基因打靶技术的概念 …… (135)
二、胚胎干细胞技术的发展 …… (136)
三、同源重组技术的发展 …… (137)
第二节　基因打靶技术在生物医学研究中的应用前景 …… (138)
一、基因打靶技术在基因功能研究中的应用 …… (138)
二、应用基因打靶技术研制人类疾病动物模型 …… (139)
三、应用基因打靶技术改良动物品系和研制动物反应器 …… (139)
第三节　研制基因打靶小鼠的实例—*Smad5* 基因剔除小鼠的研制和鉴定 …… (140)
一、靶基因 cDNA 探针的克隆 …… (140)
二、靶基因基因组 DNA 片断的筛选和克隆 …… (140)
三、基因剔除打靶载体的构建以及中靶 ES 细胞的筛选 …… (141)
四、基因剔除小鼠的获得及鉴定 …… (142)
五、基因剔除小鼠的表型分析 …… (143)
第四节　常用实验操作方法 …… (144)
一、小鼠基因组 DNA 的制备 …… (144)
二、基因组 DNA 的 PCR 分析 …… (144)
三、基因组 DNA 的 Southern 杂交分析 …… (144)
四、用原代小鼠胚胎成纤维细胞制备饲养层 …… (145)
五、小鼠胚胎干细胞的常规培养 …… (146)
六、ES 细胞的冻存和复苏 …… (147)

第十一章　转基因动物技术 …… (149)
第一节　原理和方法 …… (149)
一、转基因、转基因动物与转基因动物模型 …… (149)
二、目的基因的设计 …… (150)
三、转基因的方法 …… (150)
第二节　显微注射法的操作步骤 …… (153)
一、实验条件的准备 …… (153)
二、供体鼠和受体鼠的同步准备 …… (156)
三、受精卵的采集 …… (157)
四、受精卵的显微注射 …… (157)
五、转基因后受精卵的再植 …… (158)
第三节　转基因整合与表达的检测 …… (159)

一、基因组 DNA 的提取 …… (159)
二、外源基因的检测 …… (160)

第十二章 RNA 干扰 …… (164)
第一节 RNA 干扰的原理 …… (164)
第二节 RNA 干扰的生物功能 …… (165)
第三节 哺乳类中的 RNA 干扰 …… (165)
一、RNAi 在哺乳动物中存在的证据 …… (165)
二、哺乳动物中的 RNAi …… (165)
三、长 dsRNA 引发的 RNAi …… (166)
四、哺乳类稳定 RNAi …… (166)
第四节 RNA 干扰的实验过程 …… (166)
一、RNAi 的产生 …… (166)
二、给予途径 …… (169)
三、沉寂效应鉴定 …… (169)
四、存在问题 …… (169)
第五节 RNA 干扰在生物医学和功能基因组中的应用 …… (169)

第十三章 常用数据库及分子生物学软件 …… (171)
第一节 常用数据库 …… (173)
第二节 分子生物学常用软件 …… (175)

第二部分 生物芯片技术

第十四章 生物芯片概介 …… (179)
第一节 生物芯片产生的科技背景及在医学中应用的优势 …… (179)
一、生物芯片的定义和特点 …… (179)
二、生物芯片产生的科技背景 …… (180)
三、生物芯片在医学中的应用概况与前景 …… (181)
第二节 生物芯片相关分子生物学概念 …… (185)
一、基因 …… (185)
二、基因组和基因组的信息含量 …… (185)
三、真核基因组和原核基因组 …… (186)
四、基因表达 …… (187)
五、基因突变 …… (187)
六、基因多态性 …… (187)
七、DNA 的变性与复性 …… (188)
八、探针 …… (188)

九、杂交 …… (189)
十、PCR …… (190)
第三节　生物芯片的分类简介 …… (191)

第十五章　生物芯片的制备和检测 …… (193)
第一节　载体的表面修饰 …… (193)
一、芯片载体和载体材料 …… (193)
二、载体化学修饰 …… (195)
三、载体修饰后的固定及选择原理 …… (199)
第二节　生物芯片制作方式 …… (199)
一、微阵列芯片及其制备技术 …… (199)
二、微流控芯片及其制备技术 …… (203)
三、其他类型芯片制作方式 …… (204)
四、基因芯片和寡核苷酸芯片制备原理 …… (205)
第三节　生物芯片检测的实验流程 …… (207)
一、芯片样品的制备 …… (208)
二、杂交探针的种类和标记 …… (214)
三、核酸分子杂交和杂交微环境 …… (217)
四、DNA 微阵列操作应用实例 …… (219)
第四节　芯片杂交信号的收集及生物信息的分析 …… (232)
一、芯片杂交信号的收集及处理 …… (232)
二、生物信息的分析和提取 …… (234)

第十六章　微阵列探测中的共聚焦扫描技术 …… (243)
第一节　微阵列技术、荧光及荧光探测 …… (245)
一、微阵列探测特征 …… (245)
二、荧光基础理论 …… (246)
三、荧光探针的选择 …… (247)
四、探测设备的光学设计要求 …… (248)
第二节　样品处理 …… (253)
一、标准显微载玻片 …… (254)
二、其他规格载玻片 …… (254)
三、样品环境 …… (254)
第三节　共聚焦扫描原理 …… (254)
一、共聚焦原理 …… (254)
二、共焦微阵列扫描仪 …… (255)
三、共聚焦系统优点与缺点 …… (257)
四、共聚焦扫描实现与折衷考虑 …… (257)
第四节　共聚焦扫描仪光学设计及关键特征 …… (258)

一、波长鉴别 …… (258)
二、探测器 …… (260)
三、信号处理及仪器控制 …… (260)
四、扫描仪功能参数 …… (262)
第五节 ScanArray® 型共聚焦微阵列扫描仪 …… (264)
一、扫描仪的集成化 …… (264)
二、ScanArray® 扫描仪结构组成 …… (265)
三、ScanArray® 扫描器优点 …… (266)

第十七章 生物芯片在医学中的应用 …… (268)
第一节 生物芯片技术在基因组功能研究中的应用 …… (268)
一、寻找和发现新基因 …… (268)
二、基因表达分析 …… (269)
三、DNA 序列测定与序列间比较 …… (271)
四、突变体和多态性的检测 …… (271)
第二节 生物芯片技术用于单核苷酸多态性的检测 …… (273)
一、SNPs 的检测技术 …… (273)
二、SNPs 技术的应用 …… (283)
第三节 生物芯片技术在遗传性疾病诊断中的应用 …… (290)
一、遗传性疾病的分类 …… (290)
二、遗传性疾病与基因突变 …… (291)
三、基因突变检测技术 …… (293)
四、遗传性疾病基因芯片 …… (295)
五、遗传性疾病的基因芯片的应用 …… (301)
第四节 生物芯片技术在肿瘤研究中的应用 …… (303)
一、肿瘤基因表达谱的研究 …… (303)
二、寻找肿瘤相关基因 …… (305)
三、肿瘤分型的研究 …… (309)
四、肿瘤相关基因功能的研究 …… (312)
五、肿瘤诊断的研究 …… (314)
六、筛选抗癌药物和监测肿瘤治疗效果 …… (317)
第五节 生物芯片在肝癌及病毒性肝炎诊断中的应用 …… (319)
一、基因芯片技术在肝癌研究中的应用 …… (319)
二、基因芯片技术在病毒性肝炎及其他肝病诊断中的应用 …… (323)
第六节 生物芯片技术在微生物变异及疾病耐药性检测中的应用 …… (327)
一、生物芯片在微生物基因变异检测中的应用 …… (327)
二、生物芯片在疾病耐药性检测中的应用 …… (336)
三、前景与展望 …… (342)
第七节 生物芯片技术在药学方面的应用 …… (343)

一、生物芯片在药物靶点发现与药物作用机制研究中的应用 …… (343)
二、生物芯片在毒理学研究中的应用 …… (347)
三、生物芯片作为超高通量药物筛选平台的应用 …… (348)
四、生物芯片在药物基因组学中的应用 …… (350)
五、基因芯片技术在疾病耐药性检测中的应用 …… (351)
六、生物芯片对中药物种的鉴定 …… (351)
第八节　生物芯片技术在法医学中的应用 …… (351)
一、生物芯片在法医物证学中的应用 …… (351)
二、生物芯片在法医病理学中的应用 …… (360)
三、生物芯片在痕量物证及毒品和毒物分析中的应用 …… (366)
四、前景与展望 …… (367)

第十八章　蛋白质芯片 …… (368)
第一节　蛋白质芯片产生的背景 …… (368)
一、后基因组时代 …… (368)
二、蛋白质组学的兴起与发展 …… (369)
三、传统技术的改造和新技术的产生 …… (369)
第二节　蛋白质芯片的基本原理 …… (371)
一、蛋白质芯片的特点 …… (372)
二、蛋白质芯片的分类 …… (372)
三、蛋白质芯片制作和检测的基本流程 …… (374)
四、蛋白质芯片的功能作用 …… (377)
五、蛋白质芯片的相关技术 …… (380)
第三节　蛋白质芯片在医药学上的应用进展 …… (382)
一、发现新的疾病标志物作为临床诊断的指标 …… (383)
二、对疾病病理分子机制的研究 …… (385)
三、筛选药物靶，发现新药物，研究药理机制 …… (385)
第四节　蛋白质芯片的发展展望 …… (386)

参考文献 …… (389)
封三彩图

第一部分

基因分析技术

第一章

cDNA 文库的构建

内容提要

cDNA 文库代表某一特定组织或器官的 mRNA 编码信息。mRNA 分子非常不稳定，且难于扩增。因此，可以将 mRNA 的编码信息转变成稳定的双链 cDNA，然后将其插入能够自身复制的载体中。载体中的每一个 cDNA 分子就代表着某一特定 mRNA 的编码信息。为了获得覆盖面广、质量较高的 cDNA 文库，选择好的 cDNA 合成试剂盒是建库成败的关键。本章将对构建 cDNA 文库的原理及关键实验技术和经验体会逐一介绍。

cDNA 文库的构建是分子生物学领域的一项重要技术。cDNA 是以 mRNA 为模板，在逆转录酶的作用下，在体外被逆转录为 cDNA 第一链，再以 cDNA 为模板，由大肠杆菌 DNA 聚合酶Ⅰ合成第二链，得到双链 cDNA。由于组织或细胞的总 RNA 或 mRNA 中，含有该细胞的全部 mRNA分子，因而被合成的 cDNA 产物将是各种 mRNA 拷贝的群体。当它们与质粒重组后并转化至宿主细胞中，将得到一系列克隆群体，每个克隆仅含有一种 mRNA 信息，所有克隆的总和则包含细胞内全部 mRNA 的信息，这种克隆群体则为 cDNA 文库。应用 cDNA 文库筛选和克隆新基因，是一条快速而有效的途径。也可从 cDNA 文库中筛选出所需的目的基因，用于该基因的表达和研究。

第一节　cDNA 文库构建的基本原理

成功地构建一个 cDNA 文库，从组织或细胞的获得、mRNA 的提取等每一步都很重要，一个环节出现问题，都将降低 cDNA 文库的质量。只有构建覆盖面广、质量高的 cDNA 文库，才具有较高的应用价值。因此建议使用公司生产的试剂盒，并严格按操作说明书操作。Stratagene 公司生产的 ZAP Expression cDNA Synthesis Kit（Catalog # 200403）和 ZAP Expression cDNA Gigapack III Gold Cloning Kit（Catalog # 200451）可用于 cDNA 文库的构建。

本章应用 Stratagene 公司生产的试剂盒介绍构建 cDNA 文库的原理和具体操作步骤。ZAP Expression cDNA Synthesis Kit 应用含有 *Xho* Ⅰ酶切位点的 18 个寡聚脱氧胸苷酸[oligo(dT)$_{18}$]作接头引物。该引物可与 mRNA 的多聚脱氧腺苷酸尾[poly(A)$_n$]杂交结合，在逆转录酶（StrataScript™ reverse transcriptase）的作用下合成 cDNA 第一链。在 cDNA 的第一链合成过程中，

加入甲基化的脱氧胞苷酸(5 – methyl dCTP)以取代正常的脱氧胞苷酸(dCTP),因此就可保护cDNA分子不被某些内切酶(如 *Xho* I)切割。因此,在进行 *Xho* I 酶切时,只有接头引物上的非甲基化的 *Xho* I 酶切位点可以被切割,而 cDNA 分子中的甲基化的 *Xho* I 酶切位点则不被切割,从而保证了 cDNA 分子的完整性。

一、cDNA 的合成

于第一链反应混合物中加入核苷酸混合物和反应缓冲液,在 StrataScript 逆转录酶作用下,以 mRNA 为模板,以含 50 个碱基的寡核苷酸做逆转录引物,进行第一链的合成。

逆转录引物序列如下:

5′ – GAGAGAGAGAGAGAGAGAGAACTAGT**CTCGAG**TTTTTTTTTTTTTTTTT – 3′

"GAGA"序列　　*Xho* I 位点　　poly(dT)$_{18}$

引物中的"GAGA"序列对 *Xho* I 内切酶的识别和 poly(dT)具有保护作用。*Xho* I 限制性酶切位点可使合成的 cDNA 以正义方向插入到 ZAP 表达载体中。poly(dT)$_{18}$可以与 mRNA 模板的 3′端的 poly(A)$_n$ 配对结合。

第一链合成时加入的核苷酸混合物包括正常的 dATP, dGTP, dTTP 和 5 – 甲基化 dCTP。合成的第一链的 dCTP 都含有甲基基团,可以防止 cDNA 在以后的克隆步骤中被限制性内切酶切断。

在第二链合成过程中,RNase H 可使结合在第一链 cDNA 上的 RNA 被切割成多个片段,使之作为 DNA polymerase I 的引物。DNA polymerase I 可以"缺口 – 翻译"的方式将这些 RNA 片段翻译成第二链 cDNA。参与第二链合成的核苷酸混合物中的 dCTP 为正常的 dCTP,以降低 5 – 甲基化 dCTP 掺入到第二链的可能性。这样就保证了接头引物中的限制性酶切位点可以被相应的内切酶所切断。双链 cDNA 在 *pfu* DNA polymerase 的作用下,其末端被补平。然后在 T4 DNA ligase 的作用下,将 *EcoR* I 接头连接至 cDNA 末端。接头序列为:

5′ – OH – AATTCGGCACGAGG – 3′

3′ – GCCGTGCTCCp – 5′

这些接头由 10 和 14 个碱基的寡核苷酸组成,并与 *EcoR* I 粘性末端互补。其中 10 个碱基的核苷酸是被磷酸化的,从而使其可与 cDNA 和其他接头的平头末端相结合,而 14 个碱基的寡核苷酸是被去磷酸化的,从而阻止它与其他粘性末端结合。接头连接完成后,灭活连接酶,磷酸化 14 个碱基的寡核苷酸,使其能连接至去磷酸化的载体末端。

经 *Xho* I 酶切后,可以切除 cDNA 3′端的 *EcoR* I 接头和残留的接头 – 引物。过 Sepharose CM – 2B 胶分离柱时,可将上述二个片段去除。这样经过分离的 cDNA 再经沉淀后即可连接于 ZAP 表达载体。cDNA 合成流程图见图 1 – 1。

应用高效系统如 Gigapack III Gold 包装提取物进行 cDNA 文库的包装,用 *E. coli* 细胞株 XL1 – Blue MRF′铺盘。因为大多数的 *E. coli* 细胞株能酶解含有 5 – 甲基化的 dCTP,因此一定要用这种 McrA – McrB – 的菌株铺盘。

二、ZAP 表达载体的结构

ZAP 表达载体的结构见图 1 – 2。ZAP 表达载体可以用于真核和原核细胞的表达。ZAP 表达载体含有 12 个克隆位点,可以插入 0 ~ 12 kb 大小的外源性 DNA。这 12 个克隆位点是:*Apa* I, *BamH* I, *EcoR* I, *Hind* III, *Kpn* I, *Not* I, *Sac* I, *Sal* I, *Sma* I, *Spe* I, *Xba* I 和 *Xho* I。克隆

到 ZAP 表达载体中的外源性 DNA 可以通过环化机制脱离噬菌体载体而转变成带有卡那霉素抗性的环状 pBK-CMV 噬菌粒(phagemid)。pBK－CMV 噬菌粒的结构见图 1－3。噬菌粒实际上等同于质粒,因为它是通过噬菌体环化而来,因此命名为噬菌粒。通过环化获得的噬菌粒,非常有利于基因克隆、扩增和提取等分子生物学操作。

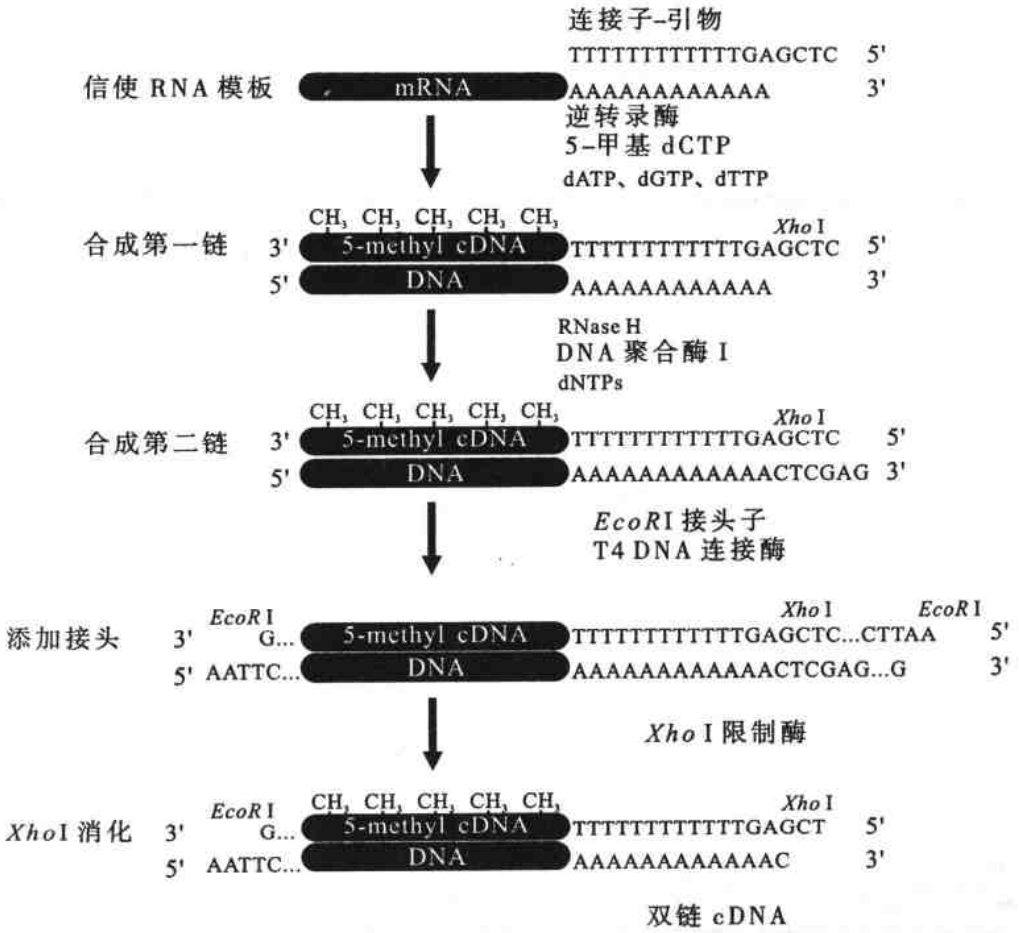


图 1－1　cDNA 合成流程图

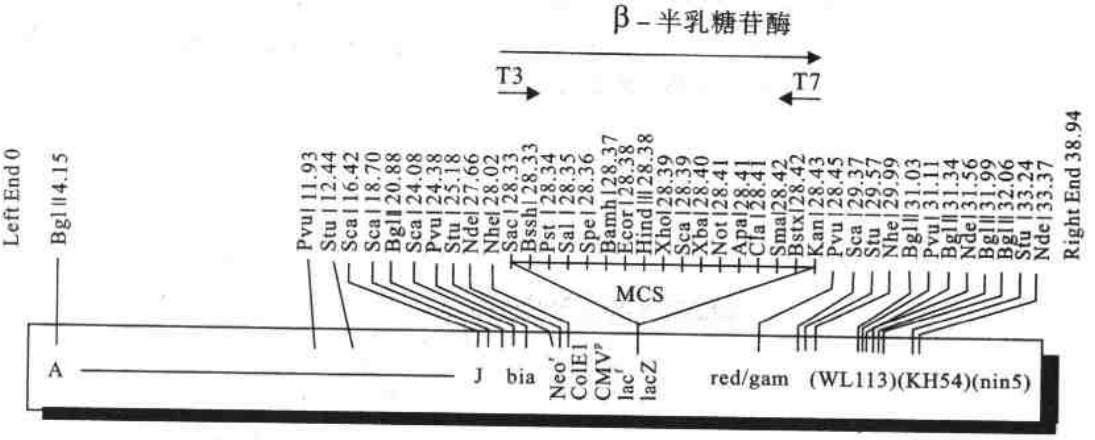


图 1－2　ZAP 表达载体的结构

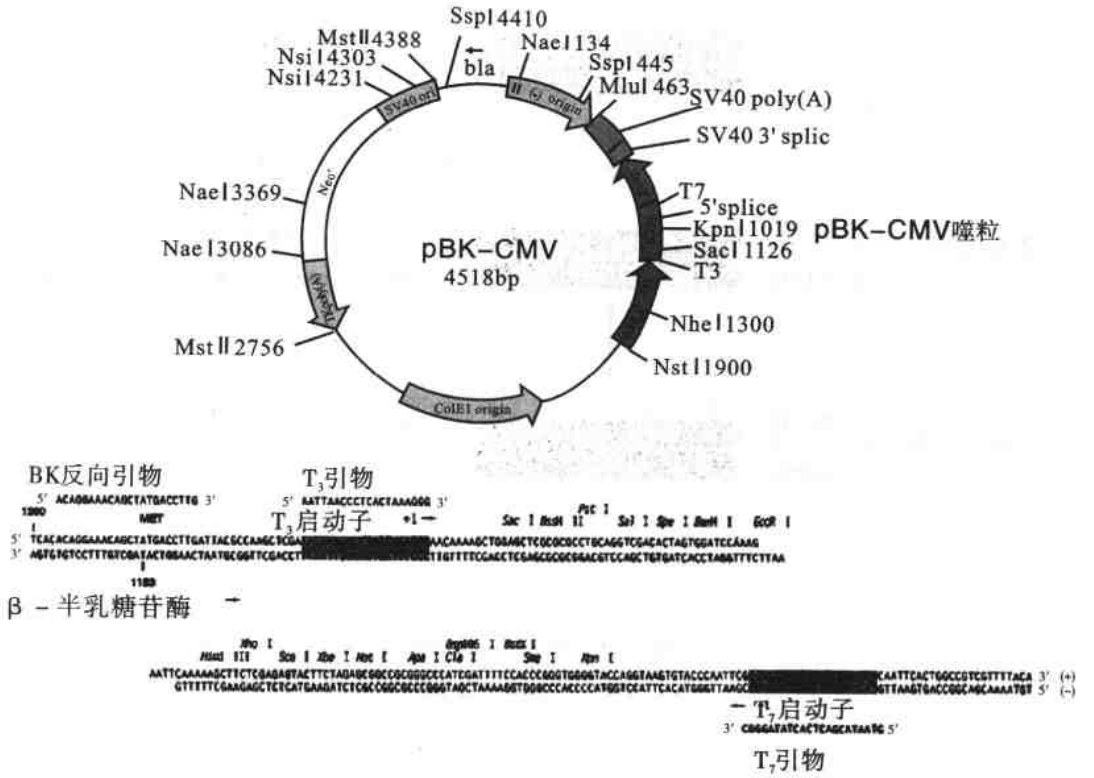


图 1－3　pBK－CMV 噬菌粒的结构

第二节　cDNA 文库构建的具体操作步骤

mRNA 的质和量是构建高质量 cDNA 文库的关键。建议使用公司生产的总 RNA 提取试剂

盒和 mRNA 纯化试剂盒提取总 RNA 和纯化 mRNA。

一、cDNA 第一链的合成

1. 准备 42 ℃水浴。

2. 冰上融化放射性同位素[α－^{32}P]dNTP(不要使用[α－^{32}P]dCTP)和除酶以外的第一链合成试剂。同位素用于第 6 步和第二链合成。非酶管短暂震荡并离心,置于冰上。

3. 第一链合成反应的终体积为 50 μL :所加反应试剂和酶的体积为 12.5 μL;所加 mRNA 和 DEPC 处理水的体积为 37.5 μL。

4. 按次序在离心管中加入下列成分:

10 倍第一链缓冲液	5 μL
第一链甲基化脱氧核苷酸混合物	3 μL
逆转录引物(1.4 μg/ μL)	2 μL
RNA 酶抑制剂(40 U/ μL)	1 μL
DEPC 处理水	X μL

5. 混合上述反应物,再加入 X μL poly(A)$^+$ RNA(5 μg),轻轻混合。

6. 室温下使引物和 mRNA 模板结合 10 min。取 0.5 μL [α－^{32}P]dNTP(800 Ci/mmol)加入另一对照管中。

7. 在第一链反应管中加入 1.5 μL StrataScript 逆转录酶(50 U/ μL)。使第一链的反应终体积达到 50 μL。

8. 轻轻混合并离心。

9. 转移第一链反应混合物 5 μL 到含有 0.5 μL [α－^{32}P]dNTP(800 Ci/mmol)的对照管中。此管作为第一链合成的质量控制管。

10. 42 ℃水浴孵育第一链反应混合物和质量控制管 1 h。

11. 为第二链合成准备 42 ℃水浴。

12. 1 h 后结束第一链合成。将含有非放射性同位素的第一链反应管置于冰上。将含有放射性同位素的第一链反应质量控制管存于－20 ℃,将用于碱性琼脂糖凝胶电泳分析。

二、cDNA 第二链的合成

1. 将短暂震荡并离心除去酶以后的第二链合成试剂置于冰上。

2. 按次序在含有 45 μL 非放射性第一链反应混合物的离心管中加入下列成分:

10 倍第二链缓冲液	20 μL
第二链脱氧核苷酸混合物	6 μL
DEPC 处理水	114 μL
[α－^{32}P]dNTP(800 Ci/mmol)	2 μL

3. 在第二链反应混合物中加入下列酶:

RNase H (1.5 U/ μL)	2 μL
DNA polymerase I (9.0 U/ μL)	11 μL

4. 轻轻弹震反应管并短暂离心,16 ℃孵育 2.5 h,不可超过 16 ℃。

5. 第二链反应结束后,立即将反应管置于冰上。

三、补平 cDNA 链末端

1. 在第二链反应管中加入下列成分：

脱氧核苷酸混合物	23 μL
克隆化的 *pfu* DNA polymerase (2.5 U/ μL)	2 μL

2. 快速弹震反应管并短暂离心，72 ℃孵育反应管 30 min。不可超过 30 min。
3. 冰上融化 3 M 的醋酸钠。
4. 反应结束后，在反应管中加入 200 μL 酚∶氯仿(1∶1)并震荡。
5. 室温下以最大速度(约 12 000 r/min)离心 2 min，将上清转移至另一个新管中。
6. 加入等体积的氯仿并震荡。
7. 室温下以最大速度(同上)离心 2 min，将上清转移至另一个新管中。
8. 加入下列成分以沉淀 cDNA：

3 M 醋酸钠	23 μL
100% 乙醇	400 μL

9. 震荡后存于 -20 ℃沉淀过夜。
10. 4 ℃以最大速度(约 12 000 r/min)离心 60 min，预先在离心管外壁上标记出沉淀将出现的部位，以免丢失。
11. 小心吸出并丢弃具有放射活性的上清，避免搅动沉淀。
12. 加入 70% 乙醇 500 μL 洗沉淀，注意不可震荡。
13. 室温下以最大速度(同上)离心 2 min，注意按离心管上标记的沉淀方向离心。
14. 吸出乙醇，真空干燥。
15. 将沉淀溶于 9 μL *EcoR* I 接头中，4 ℃孵育 30 min。
16. 转移 1 μL 第二链反应混合物到另一新管中，作为第二链反应质量控制管。
17. 碱性琼脂糖凝胶鉴定第一链和第二链的合成质量。

碱性琼脂糖凝胶可以用于鉴定合成的第一链和第二链 cDNA 的大小范围。合成的 cDNA 主要集中在 1.8 kb 左右。由于第二链 cDNA 合成时掺入的[α-^{32}P]dNTP 明显少于第一链，因此，第二链的放射强度仅为第一链的 1/10 ~ 1/20(图 1-4)。

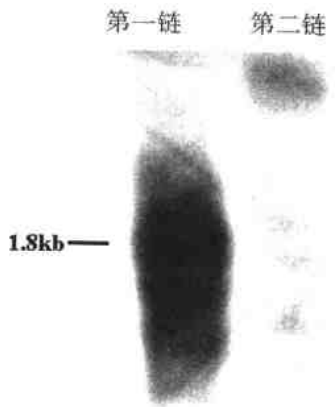


图 1-4　合成的第一链和第二链 cDNA 的大小范围

四、*EcoR* Ⅰ接头的连接

1. 在含有平端 cDNA 和 *EcoR* Ⅰ接头的管中加入下列成分：

10 倍连接缓冲液	1 μL
10 mM γATP	1 μL
T4 DNA 连接酶	1 μL

2. 短暂离心，8 ℃孵育过夜或 4 ℃连接 2 d。
3. 将反应管置于 70 ℃加热灭活连接酶 30 min。

五、*EcoR* Ⅰ末端的磷酸化

1. 将上述反应管短暂离心 2 s，室温下冷却反应管 5 min。
2. 在反应管中加入下列成分，用以磷酸化接头末端

10 倍连接酶缓冲液	1 μL
10 mM γATP	2 μL
灭菌水	5 μL
T4 多核苷酸激酶	2 μL

3. 37 ℃孵育 30 min。
4. 70 ℃加热灭活激酶 30 min。
5. 短暂离心 2 s，平衡至室温 5 min。

六、*Xho* Ⅰ酶切

1. 在上述反应管中加入下列成分：

Xho Ⅰ缓冲液	28 μL
Xho Ⅰ(40 U/ μL)	3 μL

2. 37 ℃孵育 1.5 h。
3. 在反应管中加入 5 μL 10 倍的 STE 缓冲液和 125 μL 的 100%乙醇。
4. -20 ℃沉淀反应物。
5. 沉淀后以最大速度(约 12 000 r/min)在 4 ℃离心 60 min。
6. 弃去上清，彻底干燥沉淀，将沉淀溶于 14 μL 1 倍的 STE 缓冲液中。
7. 加入 3.5 μL 柱上样染料缓冲液。此时样品即可用于过 Sepharose CL-2B 胶分离柱。

七、过柱分离 cDNA

准备分离柱和洗脱 cDNA 应在一天内完成。

(一)组装分离柱

1. 组装柱前的准备工作：

(1)从 4 ℃冰箱中取出 Sepharose CL-2B 分离胶和 10 倍的 STE 缓冲液，平衡至室温。

(2)用无菌水按 1∶10 比例稀释 10 倍的 STE 缓冲液，配制 1 倍的 STE 缓冲液 50 mL。

2. 按下列步骤组装分离柱(图 1-5 所示为组装后的分离柱)：

(1)取出塑料包装袋中无菌的 1 mL 刻度管。注意戴手套操作。

(2)用无菌的细长镊子夹一棉塞插入刻度管的尖端部，棉塞长约 3～4 mm。

(3)剪一大小合适的塑料管连接 1 mL 刻度管和 10 mL 注射器。连接好后,在刻度管和注射器之间应无缝隙。

(4)除去注射器的内芯,注射器的功能是储存洗脱缓冲液。

3. 将组装好的分离柱固定于支架上。

(二)填充分离胶

1. 按下列步骤将 Sepharose CL-2B 分离胶加入分离柱内。

(1)轻轻混匀 Sepharose CL-2B 分离胶。

(2)用玻璃吸管吸 2 mL 1 倍的 STE 缓冲液,尽可能深地插入到分离管中,加入缓冲液,注意避免气泡产生。

(3)用玻璃吸管迅速将均一的 Sepharose CL-2B 分离胶加入至分离管中,玻璃吸管尽可能深地插入到分离管中。重复加入分离胶,直至分离胶表面高度至刻度管和注射器连接处下端 1/4 英寸处。此过程中注意避免气泡产生。

2. 用至少 10 mL 1 倍的 STE 缓冲液洗柱。大约需要 2 h 完成洗柱步骤。

3. 当分离胶表面还剩约 50 μL STE 缓冲液时,快速加入 cDNA 样品至分离胶表面。

4. cDNA 样品进入分离胶后,立即沿注射器内壁加入 3 mL 1 倍 STE 缓冲液洗柱。

5. 随着 cDNA 的洗脱,染料在分离柱中逐渐弥散下移。

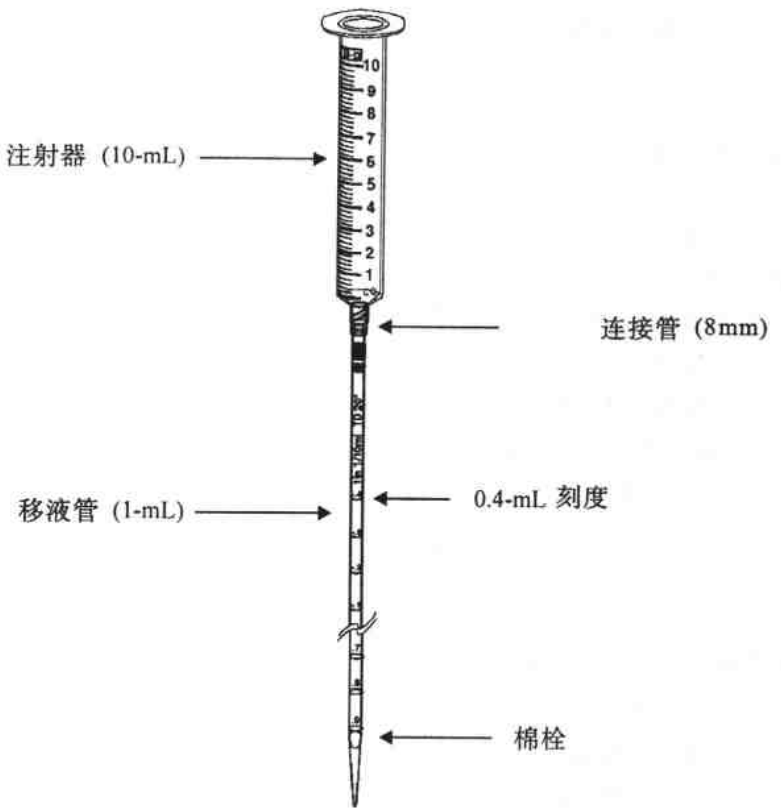


图 1-5 组装后的分离柱

(三)收集分离的 cDNA 样品

Sepharose CL－2B 分离胶可按分子量的大小分离 cDNA。分子量大的 cDNA 先流出,然后是较小的 cDNA,最后是游离的核苷酸。用便携式的同位素计数仪,在洗脱过程中一般可以检测到二个放射活性峰。第一个洗脱峰代表 cDNA。第二个洗脱峰代表未掺入的有放射性的核苷酸。

1. 当染料的前缘到达分离管 0.4 mL 刻度时开始收集 cDNA。为了获得长度在 400 bp 以上的 cDNA,需共收集 12 管,每管 3 滴(约 100 μL)。

2. 从每一收集管中吸出 8 μL,进行 5%非变性聚丙烯酰胺凝胶分析(图 1－6),以确定哪几个收集管中的 cDNA 用于连接反应。一般选取含有平均 cDNA 长度较长的几个收集管合在一起,用于下面的纯化和连接反应(如图 1－6 中的 2、3 和 4 管)。

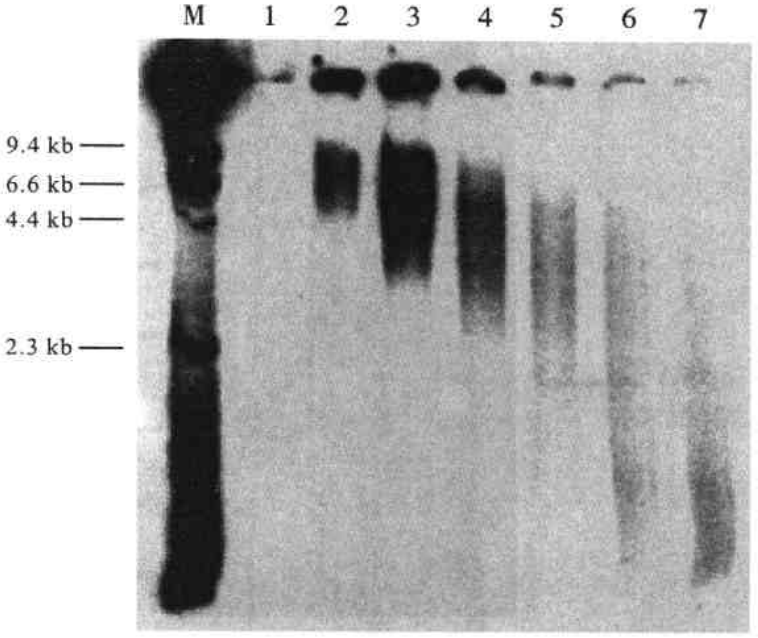


图 1－6　5%非变性聚丙烯酰胺凝胶分析各收集管中 cDNA 的大小

(四)纯化分离出来的 cDNA

酚－氯仿抽提,乙醇沉淀,回收 cDNA。本步骤主要目的是除去混杂在 cDNA 中的蛋白质,尤其是激酶。

1. 在 cDNA 管中加入等体积的酚－氯仿(1:1,V/V)。

2. 震荡,室温下离心 2 min。转移上层水相至另一新离心管中。

3. 加入等体积的氯仿。

4. 震荡,室温下离心 2 min。转移上层水相至另一新离心管中。

5. 加入 2 倍体积的 100%乙醇。

6. －20 ℃沉淀过夜。

7. 4 ℃以最大速度 (约 12 000 r/min) 离心 60 min。转移上清至另一管中。为了确保 cD-

NA 被完全回收，用便携式的同位素计数仪检测沉淀的放射活性。如果上清的放射活性很高，则需重复离心。否则弃去上清。

8. 小心用 80% 乙醇洗沉淀。不可震荡！按沉淀方向以最大速度在室温下离心 2 min。除去乙醇，室温下干燥 5 min。

9. 用便携式的同位素计数仪，进行每秒钟计数(CPS)。如果 CPS < 30，将 cDNA 沉淀溶于 3.5 μL 的水中；如果 CPS > 30，则溶于 5 μL 的水中。用加样枪头充分混合。

10. 应用含有溴化乙锭的琼脂糖凝胶盘，进行 cDNA 的定量。

取 0.5 μL cDNA 样品和不同稀释度的标准 DNA 样品点样于含有溴化乙锭的琼脂糖凝胶盘上，在紫外光下比较 cDNA 样品和不同稀释度的标准 DNA 样品亮度，即可以比较准确地对 cDNA 样品进行定量，见图 1－7。图中所示的 cDNA 样品含量大于 200 ng/ μL。

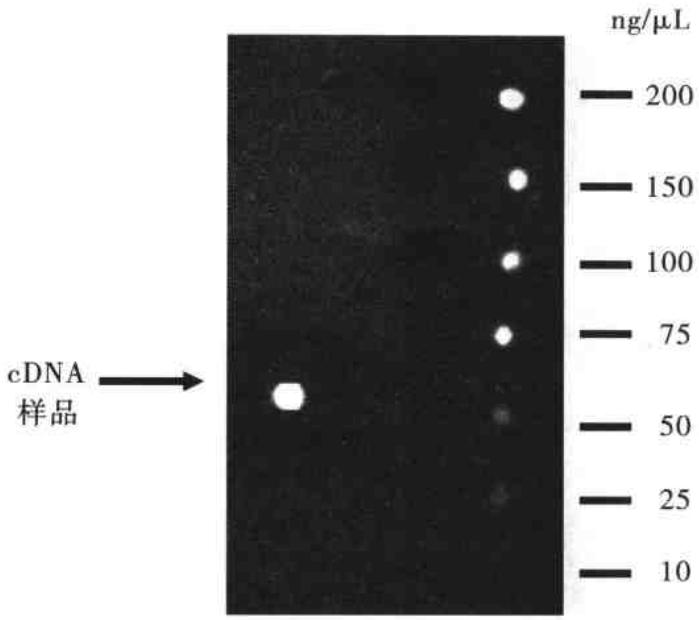


图 1－7　cDNA 的定量

八、cDNA 与 ZAP 表达载体的连接

1. 按次序在连接管中加入下列成分：

溶好的 cDNA (100 ng)	X μL
10 倍的连接酶缓冲液	0.5 μL
10 mM 的 γATP (pH 7.5)	0.5 μL
ZAP 表达载体(1 μg / μL)	1.0 μL
加水至终体积	4.5 μL

最后加入 0.5 μL 的 T4 DNA 连接酶(4 U/ μL)。

2. 4 ℃孵育 12 h 至 2 d。

3. 连接反应结束后，即可用于包装。

九、准备宿主菌

1. 从含有 XL1 - Blue MRT 菌株的甘油保存管中，用涂菌环蘸取少许菌液，划含有四环素的琼脂盘。37 ℃培养过夜。

2. 挑单菌落，加入适量的培养液中。培养液中加入 10 mM $MgSO_4$和 0.2%麦芽糖。

3. 37 ℃摇菌生长 4 ~ 6 h(生长不要超过 OD_{600} = 1)。

4. 菌液于室温 500 g 离心 10 min，弃去上清。

5. 用原菌液一半体积的 10 mM $MgSO_4$轻轻悬浮细菌。

6. 用 10 mM $MgSO_4$稀释细菌，使其 OD_{600}为 0.5。细菌保存在冰上，应在数小时之内使用。

十、应用 Gigapack III Gold 包装蛋白包装含有 cDNA 的 ZAP 重组载体

1. 从 - 80 ℃冰箱中取出包装蛋白管，置于干冰上。

2. 将包装蛋白管置于手指中间，使其快速融化。

3. 取 1 ~ 4 μL 连接反应物加入包装蛋白中。

4. 用加样枪头快速混匀，并避免气泡产生。

5. 快速离心 3 ~ 5 s。

6. 室温(22 ℃)孵育 2 h。不可超过 2 h。

7. 加 500 μL 的 SM(试剂盒提供)缓冲液。

8. 加 20 μL 的氯仿，轻轻混匀。

9. 短暂离心。

10. 上清即含有噬菌体。可以保存于 4 ℃冰箱中。

十一、铺盘和库容量鉴定

1. 准备宿主菌。

2. 按下列方法用 SM 缓冲液稀释噬菌体：进行原始库容量测定时，取 1 μL 噬菌体加到 200 μL宿主菌中(用 10 mM $MgSO_4$稀释，使其 OD_{600}为 0.5)；进行扩增库容量测定时，按下列比例稀释噬菌体：1∶10 000，1∶100 000，1∶1 000 000。各取 1 μL 稀释后的噬菌体加到 200 μL 宿主菌中(用 10 mM $MgSO_4$稀释，使其 OD_{600}为 0.5)。

3. 37 ℃孵育细菌和噬菌体混合物 15 min，使噬菌体黏附于细菌上。

4. 加入下列成分：

2 ~ 3 mL NZY 上层琼脂(融化，冷却至 48 ℃)

15 μL IPTG(0.5 M)

50 μL X - *gal*(250 *mg*/mL)

5. 混匀后迅速铺 NZY 上层琼脂盘。10 min 后倒置琼脂盘，37 ℃孵育 6 ~ 8 h 可见噬菌斑长出。

6. 非重组的噬菌斑呈蓝色，而重组的噬菌斑呈白色。重组的噬菌斑的数目应是非重组的噬菌斑的 10 ~ 100 倍。

原始库不稳定，因此，应尽快扩增原始库后再保存。

十二、原始库的扩增

通过扩增 ZAP 表达原始 cDNA 文库，可以获得大量而稳定的 cDNA 文库。最好不要进行

多次扩增,否则将会丢失生长较慢的克隆。

第一天:

1.准备宿主菌。

第二天:

2. 用 10 mM $MgSO_4$稀释宿主菌,使其 OD_{600}为 0.5。细菌保存在冰上,每个 150 mm(直径)盘使用宿主菌 600 μL。

3. 根据原始库容量的大小,将包装混合物分成若干份,使每份含有 5×10^4个噬菌斑。将每一份与 600 μL 的宿主菌混合。如要扩增 1×10^6个噬菌斑,需将原始库分成 20 等份,每份含有 5×10^4个噬菌斑,铺 20 个 150 mm NZY 琼脂盘 。

注意:每份包装混合物的体积不可超过 300 μL/600 μL 宿主菌。

4. 37 ℃孵育细菌和噬菌体混合物 15 min。

5. 将 6.5 mL NZY 上层琼脂(融化,冷却至 48 ℃)与上述混合物混合,平铺于 150 mm NZY 琼脂盘上。

6.NZY 上层琼脂凝固后,倒置于 37 ℃孵育 6 ~ 8 h。

7. 向盘内加入 8 ~ 10 mL 的 SM 缓冲液。保存在 4 ℃,并在摇床上轻轻摇晃,使噬菌体弥散至 SM 缓冲液中。

第三天:

8. 从每个盘中回收悬浮液,合并在另一无菌的离心管中。每个盘再用 2 mL 的 SM 缓冲液冲洗,一并转入上述离心管中。加入 5% 氯仿。混合后置室温 15 min。

9. 以 500 g 离心 10 min,以除去细菌残渣。

10. 回收并转移上清至另一无菌离心管中。如果上清浑浊或含有较多的细菌,则重复离心。如果上清清亮,则加入 0.3% 氯仿,并存放于 4 ℃。最好将扩增库分成若干份,加入 7% DMSO,可长期保存于 - 80 ℃。

11. 用宿主菌和不同稀释度的扩增库,测定扩增库的容量。扩增库的容量应为 $10^9 \sim 10^{11}$ 噬菌斑/mL。

计算公式如下:

$$\frac{\text{噬菌斑数} \times \text{稀释倍数}}{\text{铺盘量}} \times 1\,000\ \mu\text{L/mL}$$

注:构建 cDNA 文库,除了本文介绍的方法外,还有其他方法,读者可自行选择。

(魏英杰)

第二章

差减杂交法和差减文件的构建

内容提要

差减杂交法是基于DNA分子的序列差异或mRNA(cDNA)分子的拷贝数差异通过分子杂交过程分离出差异性的DNA或cDNA片段，它是一种原理简单，然而实用有效的分离相关基因的方法，特别适用于一般实验室的基础性研究。虽然它的操作相对复杂，但它从二三十年前最初尝试直到现在还在不断改进和广泛采用，是分子生物学技术长期发展的见证之一。本章主要介绍了各种常用的差减杂交法的基本特点和大致流程包括差减文库的构建，并附有两种主要方法的范例。

差减杂交法是目前广泛应用于差异性基因分离和鉴定的方法之一。简单地讲，差减杂交法是基于不同样品间特定核酸序列(基因组DNA或mRNA)在数量(等位基因数目或拷贝数或表达量)上的差异，通过分子间的同源性杂交将存在数量差异性的序列分离出来。从研究对象和使用的基本材料来看，大致可以分为两类，即基因组差减杂交法和mRNA或cDNA差减杂交法。基因组差减杂交法使用基因组DNA作为研究对象和实验材料，主要应用于肿瘤特异性缺失片段的筛选和各种遗传性疾病的家系研究以及进化生物学的研究；mRNA或cDNA差减杂交法则以基因转录产物mRNA或其逆转录产物cDNA作为研究对象和实验材料，是最早开发和应用于基因表达差异分析的经典方法之一，已经和正在广泛应用于包括生物学和医学的各个领域。

第一节　基因组差减杂交法

遗传学上的多样性决定了每个个体间在基因组序列上的差异性(同卵双生除外)，同时也决定了个体间在生物学特性上的差异。这种差异性的分子本质即遗传物质的基础一直就是科学研究所关注的焦点，如已知各种肿瘤发生中伴随的基因组DNA缺失或易位，可传递性外源遗传物质所介导的各种生物毒性和抗性等。基因组差减杂交法提供了一种分析类似问题的实用工具。简单地讲，其基本操作就是将不同来源的基因组DNA用特定的限制性内切酶进行消化(通常使用识别四位点的酶切，可以酶切得更完全和形成相对短的片段以利于完全的杂交反应)，其中一种样品来源的DNA过量并进行生物素标记，将两种样品混合杂交，一起变性和复性，然后通过生物素和亲和素亲和性结合的特点，去除生物素标记样品来源的DNA以及杂交

双链，剩余的 DNA 中加入新的过量的并标记的异源 DNA 样品，重复进行几次杂交反应，达到最大限度地驱除两者共有序列的目的，最后得到的单一样品来源的双链 DNA，并通过特异性的酶切末端连入适当的载体构建差减文库，进行进一步的鉴定。鉴定的方法包括以标记的不同样品来源的原始 DNA 为探针，对差减文库进行菌落原位杂交或斑点杂交，或者以差减文库的 DNA 片段作为探针，对不同样品来源的原始 DNA 进行 Southern 杂交或细胞和组织原位杂交，然后进行序列分析(简单的操作流程见图 2－1)。

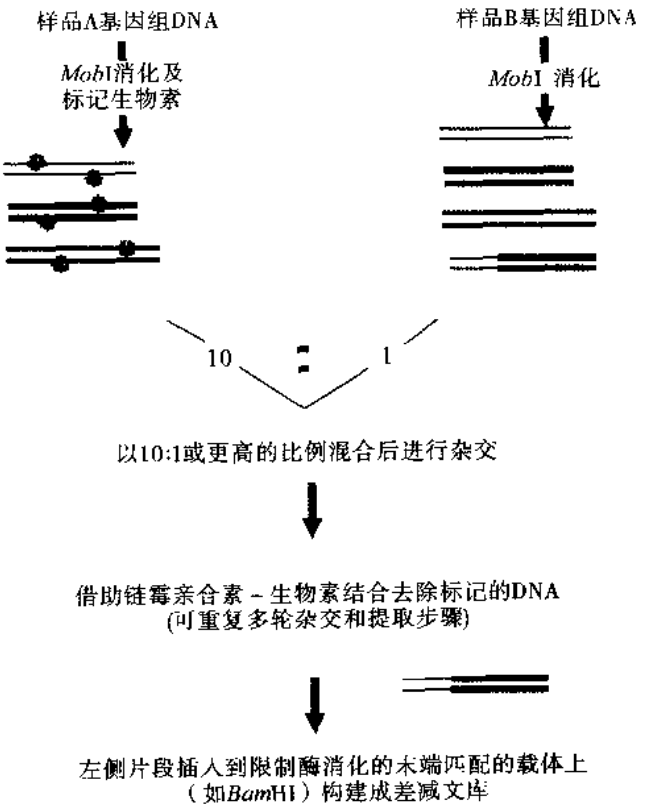


图 2－1　基因组 DNA 差减杂交操作流程示意图

在形态学检查中不能分辨的染色体或 DNA 变异性，通过控制杂交条件，可以分离出那些存在细微变化的染色体片段或基因序列。本方法的缺点是工作量大，技术要求高，同时需要大量的 DNA 样品，对于取材困难和来源少的一些病理组织，对实验设计和实际操作的要求就更高一些。经过不断的改进和其他新方法的引入，基因组差减杂交法将会越来越成熟并得到更广泛的应用。

第二节　mRNA 或 cDNA 差减杂交法

mRNA 或 cDNA 差减杂交法的分析对象则是基因的活性表达形式即信使 RNA(mRNA)以及其体外加工形式(cDNA)。由于任何基因功能的体现都是以其激活表达为前提的，因此了解各种基因的时序性或区域性的特异性激活表达成为开启生命奥秘的钥匙，这也是目前基因研究中最重要和时刻面临的问题。mRNA 或 cDNA 差减杂交法的最初设想来源于对生物发育过

程的兴趣。在正常发育的不同阶段，细胞有序地关闭或启动不同的基因表达活性，利用不同发育时期的生物材料进行差减杂交以找到那些程序性表达的基因成为发育生物学家们必须解决的方法学难题。最原始的差减杂交法是将一种材料的单链 cDNA 与过量的另一材料来源的 mRNA 进行杂交，然后去除杂交双链和 mRNA，留下未杂交的 cDNA，进一步合成双链 cDNA，并进行克隆和序列分析，从而了解发育过程中基因表达变化的情况。由于通常情况下许多生物学特性取决于各种基因表达水平变化的协同作用，针对单一基因变化的研究就显出很大的局限性，只有全面掌握各种基因的不同变化甚至是动态变化，对于了解基因型和表型的互动关系尤为重要，而 mRNA 或 cDNA 差减杂交法正是这样一种有益的工具。正因如此，采用这一方法取得的研究结果也越来越多地出现在各种专业杂志上。和所有其他方法一样，差减杂交法从最初的尝试到成为常用的研究工具经历了一个不断改进的过程。改进的重点在于如何提高差减的效率以及方法学的简易性，从开始采用简单的柱层析到现在越来越多使用的抑制性差减杂交(Suppression Subtractive Hybridization，SSH)。其他领域的方法引入也起到了有力的推动作用。实际上许多过去使用过的方法经过适当的改进目前仍然在不同的实验室采用。所有方法各有优劣，不同的研究者有不同的取舍方式。下面就按照差减杂交法大致的发展历程介绍几种常用的方法。

一、羟基磷灰石(HAP)柱层析法

这是最早运用于 cDNA 差减杂交的技术手段。它利用羟基磷灰石吸附双链核酸分子而滤过单链分子的特点，将单链和双链分子分离。对于 mRNA 和 cDNA 间的异源杂交分子，杂交双方的分子形式均为单链 mRNA 和单链 cDNA，经过变性和复性的杂交过程之后，双链的异源杂交分子与剩余的单链分子经过柱层析分离开来。mRNA 以碱水解或酶消化去除，经过几次重复杂交和分离之后，最后剩余的单链 cDNA 分子即为差异性的基因表达产物(简单的操作流程见图 2-2)。这一方法原理非常简单，实验耗费低，普通的实验室即有条件实施。其缺点在于操作中需要大量的材料(mRNA 及合成的单链 cDNA)。mRNA 分子的不稳定性决定了操作过程必须非常小心严格，难度较大。这是早期的差减杂交法，限于技术局限所通常采用的手段，但是对于经费相对缺乏的实验室只要规范操作，该方法仍值得尝试。现在有一种更新的方法是以各种固相载体如琼脂糖颗粒或磁珠共价结合寡核苷酸引物[oligo(dT)]，通过逆转录将单链 cDNA 固定在固相载体上，以这一载体作为 Driver 与另一样品来源的 mRNA(Tester)反复进行杂交和洗脱，达到去除双方共有的基因片段的目的。剩余的 mRNA 通过逆转录合成 cDNA，插入适当的载体构建差减文库，进行进一步的鉴定和序列分析。由于固相载体结合的 Driver cDNA 可以重复利用，使操作更为方便，同时也能节省许多材料。但是这一方法同样是对 mRNA 进行直接操作，技术要求高，这也是本方法实施的困难之处。

二、生物素－链亲和素的亲和分离

生物素－链亲和素的亲和分离是对羟基磷灰石柱层析法的改进。它根据生物素和亲和素稳定结合的特点，以生物素标记 DNA 双链，然后通过亲和素去除标记的 DNA。生物素标记大致有 3 种方式：①生物素标记的 dUTP 通过逆转录或 PCR 扩增反应掺入 DNA 双链；②人工合成的生物素标记的引物通过逆转录或 PCR 扩增反应掺入 DNA 双链；③光敏生物素通过强光照射激活成活性形式与碱基中的伯胺基团共价结合，从而进行标记。亲和素标记链的分离则用最简单的酚抽提法、亲和素柱层析法和亲和素磁珠法。体外合成标记和亲和分离的简单操作流

程(简单的操作流程见图 2－3)为:两种(组)不同来源的样品总 RNA 或 mRNA(通常分别命名为 Tester 和 Driver)经过逆转录合成 cDNA 第一条链和互补双链,末端补平或限制性内切酶处理后连接上不同的接头(Adaptor),通过生物素标记引物进行扩增反应或使用光敏生物素直接进行光学交联对其中的 Driver cDNA 进行标记。将过量的标记 Driver cDNA 和适量的 Tester cDNA (10:1 或更高的比例)混合进行杂交反应。杂交产物通过生物素－链霉亲和素的亲和特性进行分离(柱分离或酚抽提)从而去除 Tester cDNA 与 Driver cDNA 中重复表达的基因片段,分离后剩余的产物可以通过扩增反应增加。一个理想的差减杂交结果往往需要重复 3 次或更多次的杂交过程以达到足够高的富集效率。最后一次杂交后的剩余产物经过适当的酶切(Adaptor 序列中的位点)后可插入适当的载体构建差减文库,进行进一步鉴定和序列分析。这一方法的优点在于对低丰度表达的基因的富集效率极高,同时可操作性强;缺点在于得到的目的片段往往较短,同时差减双方共有的片段会混入文库造成一定比例的假阳性,需要进一步鉴定和筛选。

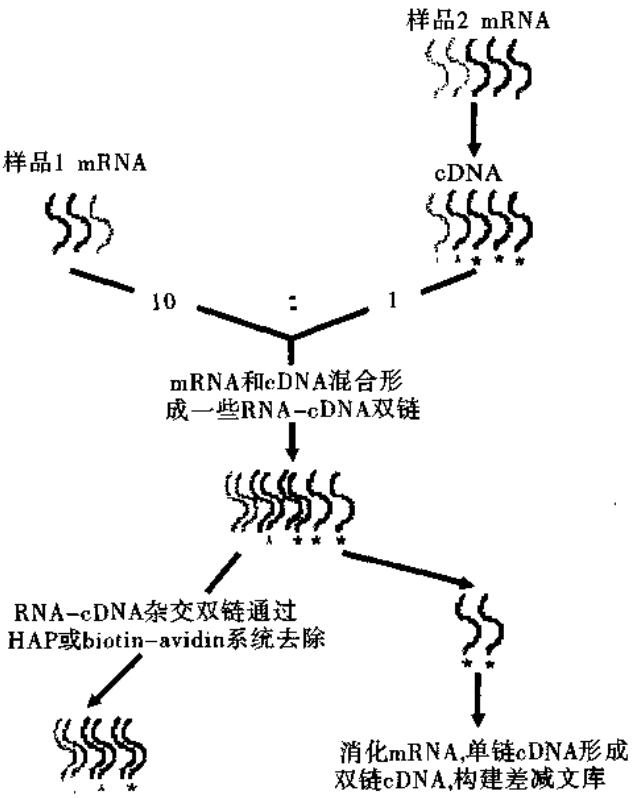


图 2－2　mRNA－cDNA 差减杂交法操作流程示意图

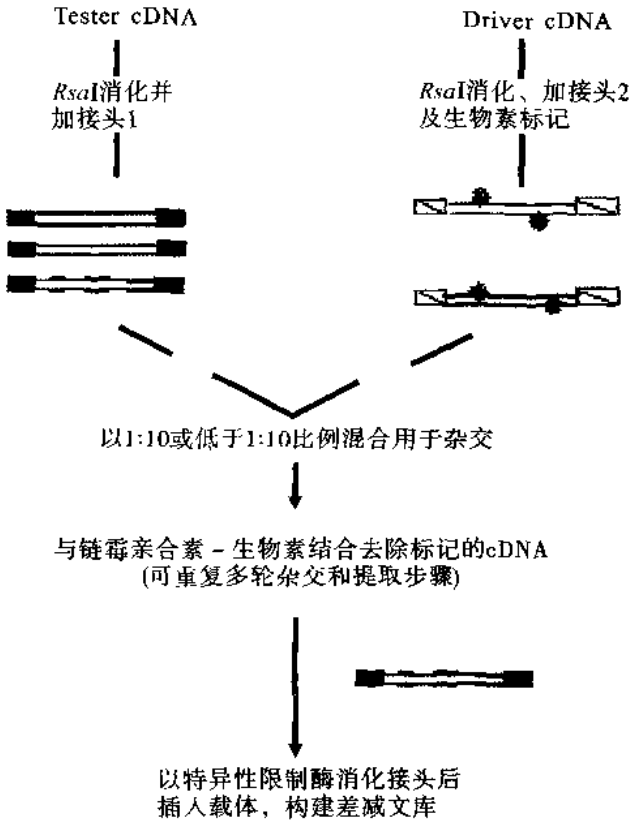


图 2－3　cDNA 差减杂交操作流程示意图

〔范例 1〕cDNA 差减杂交法分离大鼠压力负荷型心肌肥厚相关基因。

1. 大鼠压力负荷型心肌肥厚模型的建立：常规的腹主动脉缩窄法(具体操作见相关文献，此处略)。

2. 大鼠心肌细胞分离：采用常规的循环灌流分离(具体操作见相关文献，此处略)。

3. 实验组和对照组大鼠心肌细胞 RNA 提取：按常规方法(最好用商品试剂盒)。细胞离心收集后加 0.5 mL Trizol RNA 提取试剂直接匀浆(枪头反复吹打几次或小号针头抽打几次)，然后加 1/5 体积的氯仿剧烈振荡以充分混匀，静置约 10 min。4 ℃ 12 000 r/min 离心 15 min，取上清加等体积异丙醇，混匀后静置 10 min 4 ℃ 12 000 r/min 离心 10 min，去上清加 1 mL 70% 乙醇悬起沉淀，再 4 ℃ 12 000 r/min 离心 2 min，尽可能去掉上清。室温晾干。加 20 μL DEPC 无 RNA 酶水溶解(浓度在 5 μg/ μL 左右)。

4. 合成 cDNA 第一条链：

total RNA 50 μg 或 mRNA 1～2 μg	约 10 μL
Oligo$(dT)_{18}$(10 μM)	1 μL
DEPC 无 RNA 酶水加至总体积	12 μL

混匀后，70 ℃变性 5 min，随即置冰上冷却 2 min。

加入：5 × First－Strand Buffer	4 μL(试剂盒中含有)

0.1 M DTT	2 μL
dNTP Mix (10 mM)	1 μL
AMV Reverse Transcriptase (20 U/ μL)	1 μL
总体积	20 μL

混匀后短暂离心,42 ℃温育 1.5 h。

5.合成 cDNA 第二条链:

第一链反应产物	20 μL
5 × S – Strand Buffer	16 μL
dNTP Mix (10 mM)	1.6 μL
DNA Polymerase I	4 μL
DEPC 无 RNA 酶 H_2O	38.4 μL

总体积 80 μL,混匀后短暂离心,16 ℃温育 2 h。

6. 加入 2 μL T4 DNA Polymerase(4 U/ μL),混匀后 16 ℃温育 30 min。

7. 加入 100 μL 酚:氯仿:异戊醇(25:24:1),剧烈振荡混匀,4 ℃ 14 000 r/min 离心 10 min。

8. 将上清移至一支干净的 0.5 mL 离心管,加入 100 μL 氯仿:异戊醇(24:1),剧烈振荡混匀,4 ℃ 14 000 r/min 离心 10 min。

9. 取上清加入 40 μL 4 M NH_4OAC 和 300 μL 95% 乙醇,剧烈振荡混匀后 4 ℃ 14 000 r/min 离心 20 min。弃上清,加入 200 μL 70% 乙醇悬起沉淀,再 14 000 r/min 离心 5 min,尽可能移走所有上清。沉淀在室温自然干燥后直接进入连接反应。

10.cDNA 样品(包括实验组和对照组 cDNA)分别连上特殊的接头(Adaptor)。接头的序列为:

5′ CTCTTGCTTGAATTCGGACT 3′

3′ AACGAACTTAAGCCTGA 5′(有横线者为 *EcoR* I 位点)

接头连接反应如下:

实验组和对照组 cDNA 加灭菌水 5 μL

Adaptor (10μM)	2 μL
5 × Ligation Buffer	2 μL
T4 DNA Ligase (400 U/ μL)	1 μL
总体积	10 μL

混匀后短暂离心,16 ℃过夜(12 ~ 20 h)连接。反应结束后过柱(Sephacryl 400),大于 300 bp 的片段被收集,样品保存于 – 20 ℃。

11.光敏生物素标记:取部分 cDNA(包括实验组和对照组)分别标上生物素。

在 0.5 mL 离心管中加入:

实验组或对照组 cDNA	8 μL
光敏生物素(10 μg/ μL)	2 μL
加灭菌水至总体积	50 μL

将离心管置于冰上,1 000 W 碘钨灯在离液面 10 cm 处直接照射,每次 10 min,连续 3 次。标记产物以正丁醇反复抽提 3 次直至抽提液无可见红色。

12. 杂交反应:将生物素标记的实验组或对照组 cDNA 与未标记的对照组或实验组 cDNA 分别混合(比例大致为生物素标记的 cDNA:未标记的 cDNA = 10:1,cDNA 的含量比)进行杂交。

在 0.5 mL 离心管中加入:

生物素标记的实验组或对照组 cDNA	约 50 μL
未标记的对照组或实验组 cDNA	约 20 μL
10×杂交缓冲液	8 μL
加灭菌水至总体积	80 μL

98 ℃变性 2 min,然后置于 68 ℃杂交 16～20 h。

13.标记 cDNA 的去除:杂交反应结束后加入 20 μg 链霉亲和素,充分混匀后置室温 30 min,加等体积(约 100 μL)酚:氯仿,抽提 3 次直至中间相无明显沉淀物。上清转入新的 0.5 mL 离心管,加入等体积异丙醇,沉淀过夜。4 ℃ 12 000 r/min 离心 15 min,去上清,加 250 μL 70%乙醇悬起沉淀,12 000 r/min 离心 5 min,尽可能去掉上清,室温晾干,加 10 μL 灭菌水溶解。

14. PCR 扩增:杂交剩余的产物分别以接头特异的引物(5′CTCTTGCTTGAATTCGGACT 3′)进行扩增。95 ℃预变性 2 min;95 ℃ 30 s;50 ℃ 1 min;72 ℃ 1 min;30 个循环。

15. 重复差减杂交过程:以上轮两组杂交剩余的产物分别作为新的实验组和对照组cDNA,重复步骤 11～14。一个完整的差减杂交过程需要 4～5 次重复以达到最大程度富集差异性表达基因的效果。

16.差减文库的构建:几轮杂交结束后,最后剩余的产物经 PCR 扩增,以接头中含有的 *EcoR* Ⅰ位点酶切后连入载体(pBluescript－KS)构建差减文库。

酶切反应:

cDNA	10 μL
10×酶切缓冲液	2 μL
*EcoR*Ⅰ	1 μL
加灭菌水至总体积	20 μL

37 ℃温育 2 h,加灭菌水至 100 μL,然后加 100 μL 酚:氯仿和氯仿分别抽提一次。上清加入等体积异丙醇,沉淀过夜。4 ℃ 12 000 r/min 离心 15 min,去上清加 1 mL 70%乙醇悬起沉淀。再 12 000 r/min 离心 5 min,尽可能去掉上清,室温晾干,加 10 μL 灭菌水溶解。

连接反应:

cDNA 酶切产物	2 μL
pBluescript－KS(*EcoR*Ⅰ酶切并去磷酸化处理)(0.1 μg/ μL)	1 μL
5× Ligation Buffer	2 μL
灭菌水	4 μL
T4 DNA Ligase (400 U/ μL)	1 μL
总体积	10 μL

混匀后短暂离心,16 ℃过夜(12～20 h)连接。

连接产物转化 100 μL 感受态细菌(JM109)建立差减文库,同时以实验组和对照组标记的 cDNA 作为探针与差减文库进行菌落原位杂交和斑点杂交,挑选出差异性的克隆,并进一步以挑选出的差异性克隆片段为探针对原始的样品 RNA 进行 Northern Blot,以确证这些差异性基因。随后可以进行序列分析和功能分析。由于差减杂交法只是富集在未标记样品中的上调表达基因,通过改变样品标记的顺序即同时进行两套杂交就可以鉴定出相应的下调表达的基因,至此,一个完整的差减杂交过程就完成了。

本方法引入了 PCR 扩增模板技术,减少了对样本量的限制,同时多轮杂交提高了差异性基因的富集效率。但是 PCR 反应的固有缺点即优先扩增小片段的趋势使得最后得到的大都

是小的片段，同时引物二聚体的产生也会干扰多轮次 PCR 反应的效率。

三、基于 PCR 反应的差减杂交法或抑制性差减杂交(SSH)

这是目前比较常用的差减杂交法，它利用了 PCR 反应的特点，首先达到扩增样品的目的，同时通过引物序列的结构特点对 PCR 反应的效率进行调控。简单的操作流程(图 2－4)为：两种(组)不同来源样品的总 RNA 或 mRNA 经过逆转录合成 cDNA 第一条链和互补双链，经限制性内切酶 *Rsa* I(4 碱基识别位点)处理形成平末端的小片段。其中一种(组)样品(通常是 Tester cDNA)连接上不同的特殊接头(Adaptor1 和 Adaptor2)，分别与过量的 Driver cDNA 混合进行第一轮杂交，随后将两组第一轮杂交产物混合进行第二轮杂交。两轮杂交后的产物直接进行 PCR 扩增反应。由于接头序列的特殊性，同种接头之间可以互补结合形成类似锅柄的结构，使 PCR 反应的效率得到极大的抑制，同源性结合的双链不能有效扩增。单一接头的双链能以一条链作为模板进行倍数增长；无接头的双链不能扩增，而带有不同接头的异源双链通过两种接头间不同的部分序列为引物就可以克服这种抑制性效应，从而能正常地以指数增长，这就是抑制性差减杂交的关键部分。两份 Tester cDNA 经过过量 Driver cDNA 在第一步杂交反应中筛选后，差异性表达的基因将更多以单链形式在第二步杂交反应中形成异源双链，而抑制性 PCR 反应的特异性扩增使得这部分异源双链得到富集，然后就可以建立差减文库，同时进行进一步鉴定和序列分析。

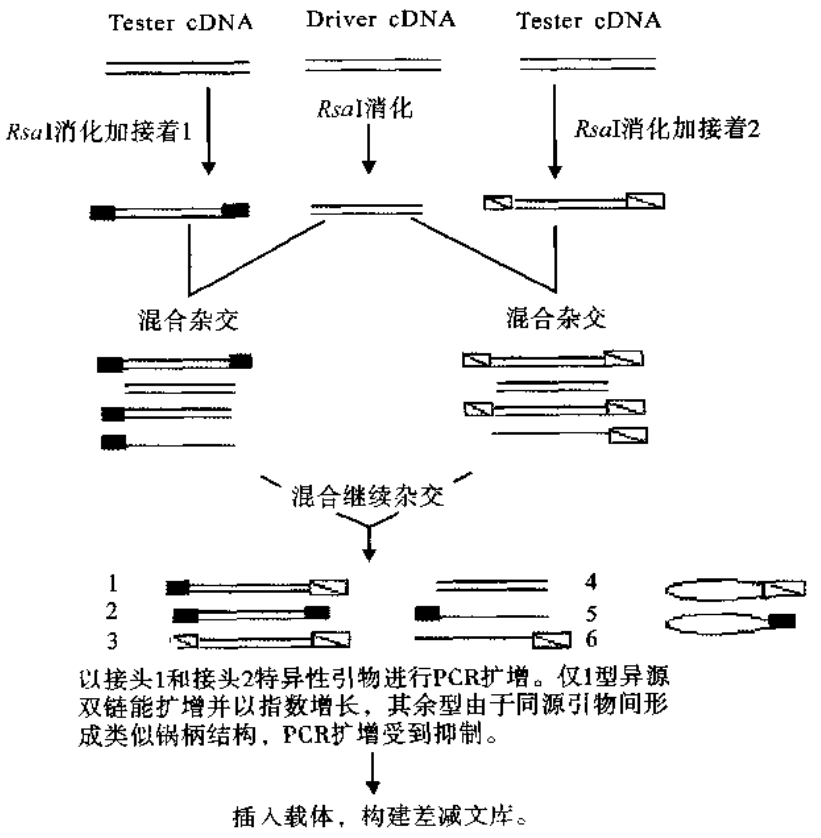


图 2－4　抑制性差减杂交操作流程示意图

〔范例 2〕应用 Clontech PCR – Select™ cDNA Subtraction Kit 进行差减杂交。

1. RNA 分离：

按常规方法。细胞培养物离心收集后加 Trizol RNA 提取试剂(1 mL Trizol/100 mm Plate)直接匀浆(用枪头反复吹打几次或小号针头抽打几次)。组织 RNA 分离则需先将组织剪切成 1 mm^3的小块，加 Trizol (0.5 mL Trizol/1g 组织) 用组织匀浆器匀浆。然后加 1/5 体积的氯仿剧烈振荡以充分混匀，静置约 10 min，4 ℃ 12 000 r/min 离心 15 min，取上清加等体积异丙醇，混匀后静置 10 min，再 4 ℃ 12 000 r/min 离心 10 min，去上清加 1 mL 70% 乙醇悬起沉淀。12 000 r/min 离心 2 min 后，尽可能去掉上清，室温晾干，加适量 DEPC 水溶解(浓度在 2 ~ 5 μg/ μL)。如果条件允许可以过亲和柱分离 Poly(A)$^+$ RNA。

2. 合成 cDNA 第一条链：

Poly(A)$^+$ RNA (2 μg) 或 total RNA(25 μg)	4 μL
cDNA Synthesis Primer (10 μM)	1 μL

总体积 5 μL，混匀后 70 ℃变性 5 min，随即置冰上冷却 2 min。

加入：5 × First – Strand Buffer	2 μL
dNTP Mix (10 mM)	1 μL
DEPC 无 RNA 酶 H_2O	1 μL
AMV Reverse Transcriptase (20 U/ μL)	1 μL

总体积 10 μL，混匀后短暂离心，42 ℃温育 1.5 h。

3. 合成 cDNA 第二链：

第一链反应产物	10 μL
DEPC 无 RNA 酶 H_2O	48.4 μL
5 × S – Strand Buffer	16 μL
dNTP Mix (10 mM)	1.6 μL
20 × S – Strand Enzyme Cocktail	4 μL

总体积 80 μL，混匀后短暂离心，16 ℃温育 2 h。

(1)加入 2 μL T4 DNA Polymerase(4U/ μL)，混匀后 16 ℃温育 30 min。

(2)加入 4 μL 20 × EDTA/Glycogen Mix 终止反应。

(3)加入 100 μL 酚：氯仿：异戊醇(25:24:1)，剧烈振荡混匀，4 ℃ 14 000 r/min 离心 10 min。

(4)小心地将上清移至一支干净的 0.5 mL 离心管，加入 100 μL 氯仿：异戊醇(24:1)，剧烈振荡混匀，4 ℃ 14 000 r/min 离心 10 min。

(5)取上清加入 40 μL 4 M NH_4OAC 和 300 μL 95% 乙醇，剧烈振荡混匀后室温 14 000 r/min 离心 20 min，弃上清，加入 200 μL 80% 乙醇悬起沉淀，再 14 000 r/min 离心 5 min，尽可能移走所有上清，沉淀在室温自然干燥 5 ~ 10 min。

(6)加入 50 μL DEPC 无 RNA 酶 H_2O 溶解沉淀。

4. *Rsa* I 酶切

ds cDNA	43.5 μL
10 × *Rsa* I Restriction Buffer	5 μL
Rsa I (10 U/ μL)	1.5 μL
总体积	50 μL

混匀后短暂离心，37 ℃温育 1.5 h。

(1)加入 2.5 μL 20× EDTA/Glycogen Mix 终止反应。

(2)加入 50 μL 酚:氯仿:异戊醇(25:24:1),剧烈振荡混匀,4 ℃ 14 000 r/min 离心 10 min。

(3)将上清移至一支干净的 0.5 mL 离心管,加入 50 μL 氯仿:异戊醇(24:1),剧烈振荡混匀,14 000 r/min 离心 10 min。

(4)取上清加入 25 μL 4 M NH_4OAC 和 18.7 μL 95% 乙醇,剧烈振荡混匀后室温 14 000 r/min 离心 20 min,弃上清,加入 200 μL 80% 乙醇悬起沉淀,再 14 000 r/min 离心 5 min,尽可能移走所有上清,沉淀在室温自然干燥 5~10 min。

(5)加入 5.5 μL 灭菌水溶解沉淀,置 -20 ℃保存。

5.所有的 cDNA 样品(包括 Tester 和 Driver cDNA)都采用上述同样的操作。下面就要将 Tester cDNA 分为两组分别连上特殊的接头(Adaptor1 和 Adaptor 2R),Driver cDNA 不需要连接接头。接头的序列分别为:

Adaptor1: Nested primer 1

5′ TAATACGACTCACTATAGGGC *TCGAGCGGCCGCCCGGGCAGGT* 3′

PCR primer 1 3′ GCCCGTCCA 5′

Adaptor 2R: Nested primer 2R

5′ TAATACGACTCACTATAGGGC *AGCGTGGTCGCGGCCGAGGT* 3′

PCR primer 1 3′ CCGGCTCCA 5′

接头连接反应如下:

Tester 1 或 Tester 2 cDNA	2 μL
Adaptor 1 或 Adaptor 2R (10μM)	2 μL
Sterile H_2O	3 μL
5× Ligation Buffer	2 μL
T4 DNA Ligase (400 units/ μL)	1 μL
总体积	10 μL

混匀后短暂离心,16 ℃过夜(12~20 h)连接。反应结束后加入 1 μL EDTA/Glycogen Mix 终止反应,72 ℃水浴 5 min 灭活连接酶。离心后样品置于 -20 ℃保存。

6.杂交反应:在两个 0.5 mL 离心管中分别加入 Driver cDNA 和连接后的 Tester 1 cDNA 或 Driver cDNA 和连接后的 Tester 2 cDNA 以及杂交缓冲液。

反应 1:

Rsa I – digested driver cDNA	1.5 μL
Adaptor 1 – ligated Tester 1	1.5 μL
4× Hybridization Buffer	1.0 μL
总体积	4.0 μL

反应 2:

Rsa I – digested driver cDNA	1.5 μL
Adaptor 2R – ligated Tester 2	1.5 μL
4× Hybridization Buffer	1.0 μL
总体积	4.0 μL

反应物加上矿物油,98 ℃变性 2 min,然后置于 68 ℃温育 8(6 ~ 12) h。

第一次杂交反应结束后立刻进行第二步杂交反应。首先准备新的变性的 driver cDNA:

Driver cDNA	1 μL
4 × Hybridization Buffer	1 μL
无菌 H_2O	2 μL
总体积	4 μL

加上一滴矿物油,98 ℃变性 2 min。立刻在一个反应管中混匀新变性的 driver cDNA 和反应 1 和反应 2 的产物(注意取样和加样动作要快,尽量避免吸进矿物油,用枪头反复吹吸几次以混匀混合物,不要加热变性!),必要时可以短暂离心。混合物置于 68 ℃温育 12 ~ 16 h。杂交结束后加入 200 μL 稀释缓冲液(pH 8.3)。

稀释缓冲液的配制:	20.0 mM	HEPES (pH 6.6)
	20.0 mM	NaCl
	0.2 mM	EDTA (pH 8.0)

充分混匀后,68 ℃加热 7 min,样品保存于 - 20 ℃。

7. 抑制性 PCR 扩增反应(suppression PCR):

第一次 PCR 反应:

稀释后的杂交产物	1.0 μL
10 × PCR reaction buffer	2.5 μL
dNTP Mix (10 mM)	0.5 μL
PCR Primer 1 (10 μM)	1.0 μL
50 × Advantage cDNA Polymerase Mix	0.5 μL
无菌 H_2O	19.5 μL
总体积	25.0 μL

混匀后短暂离心,75 ℃放置 5 min 使接头末端补平,立刻进行扩增反应。

若应用 Perkin - Elmer DNA Thermal Cycler 480 型扩增仪,其扩增条件为:94 ℃ 30 s;66 ℃ 30 s;72 ℃ 1.5 min,共 27 个循环。

若应用 Perkin - Elmer GeneAmp PCR Systems 2400 或 9600 型扩增仪,其扩增条件为: 94 ℃ 25 s 后,再 94 ℃ 10 s;66 ℃ 30 s;72 ℃ 1.5 min,共 27 个循环。

第二次 PCR 反应:

10 倍稀释的第一次扩增反应产物	1.0 μL
10 × PCR reaction buffer	2.5 μL
Nested PCR primer 1(10 μM)	1.0 μL
Nested PCR primer 2R (10 μM)	1.0 μL
dNTP Mix (10 mM)	0.5 μL
50 × Advantage cDNA Polymerase Mix	0.5 μL
无菌 H_2O	18.5 μL
总体积	25.0 μL

混匀后短暂离心,立刻进行扩增反应。

若应用 Perkin - Elmer DNA Thermal Cycler 480 型扩增仪,其扩增条件为:94 ℃ 30 s;68 ℃ 30 s;72 ℃ 1.5 min,共 10 ~ 12 个循环。

若应用 Perkin - Elmer GeneAmp PCR Systems 2400 或 9600 型扩增仪，其扩增条件为：

94 ℃ 10 s；68 ℃ 30 s；72 ℃ 1.5 min，共 10～12 个循环。

反应结束后置于 - 20 ℃保存。

由于接头和引物(PCR primer 1)序列的特殊性，第一次扩增的效率可能很低，但是会特异性地优先扩增长片段(因为短的片段受抑制性效应的影响更为显著)，从而克服一般 PCR 反应优先扩增短片段的缺点。第二次 PCR 扩增由于采用了接头内部序列(nested primer 1 and 2R)，差异性的基因将在这一步反应中得到最大程度的富集。

8. 差减文库的构建：经过两次 PCR 扩增反应后，在 Tester cDNA 中差异性表达的基因已经被有效地富集，可以将 PCR 产物直接连入克隆载体(pGEM - T - vector)构建差减文库，同时以标记的 Tester cDNA 和 Driver cDNA 作为探针与差减文库进行菌落原位杂交或斑点杂交，挑选出差异性的克隆，进一步以 Northern Blot 确证这些差异性基因，随后可以进行序列分析和功能分析。一次抑制性差减杂交只能鉴定出在 Tester cDNA 上调表达的基因，通过改变 Tester cDNA 和 Driver cDNA 的样品来源顺序即同时进行两套杂交就可以鉴定出相应的下调表达的基因，至此，一个完整的抑制性差减杂交过程就完成了。

9. 建议：首次使用该试剂盒，在合成第一链时如有条件应参照手册加入 10 微居的 dCTP 以监测 cDNA 合成及随后的 *Rsa* I 酶切。

10. 抑制性差减杂交试验的完成只是一个更大更艰巨的项目的开始。无论是差减文库还是差异克隆的分析，都是在为后续的功能试验提供动态的线索。研究者需在国际基因库和文献中发现并追踪这些线索。然后在自己的功能试验系统中进行验证和发现。

四、基因或表达序列文库(cDNA or EST Libraries)杂交

这一方法也被称作代表性差异分析(representational difference analysis，RDA)。实际上并不是一种真正意义上的杂交反应。这是一项极其费时费力的工作，它要求根据不同的材料来源建立相对完整的基因文库或表达序列文库，然后对文库进行随机测序，并将所有的结果进行统计和归一化(normalization)，即以每一种基因序列出现的次数与测序结果总数相除得到该基因的表达频率(或代表性)。由于这是一种完全数值化的方法，它的准确性取决于被测序列的总数，所测序列越多，计算出的基因表达频率约接近于真实值。另一个可以作为评价数据质量的标准是一些已知的看家基因(housekeeping genes，大多数情况下它们的表达频率被认为是比较稳定的)，这些内标准的数值稳定性也是归一化处理的基础。在所有文库进行数字化处理后，通过简单的数学分析方法可以比较不同文库中各种基因的表达变化，同时以数字化的形式表现出来。

随着人类基因组计划的飞速发展，同时推动了人们对基因文库或表达序列文库的进一步要求，毕竟基因或其表达产物才是我们真正感兴趣的话题。目前在 NCBI 已经建立了专门的主页(serial analysis of gene expression)用于收集那些公开发布的各种基因表达的文库资料，同时还有用于比较不同文库差异的应用工具(X - profiler or library profiler)。可以选择实验者感兴趣的各种文库资料进行计算机分析。随着这一数据库的不断扩充和完善，将来在世界上不同的实验室可以同时进行各种类似的工作而省去大量重复而烦琐的实验操作过程。

除了繁重的工作量外，这种方法的最大问题在于基因文库的完整性和处理数据的可靠性。首先是基因文库很难涵盖所有的表达基因，特别是低丰度的基因；其次是随机测序的数量是否满足统计学的要求(样本量、可信度等)；最后是所测序列片段的代表性。由于基因家族和同源

性序列以及许多未知基因的存在，相同的一段测序结果（通常是3′端的几百个碱基）可能来源于不同的基因，从而导致重复记数或误算。

最终了解人类自身和周围世界上形形色色的生命现象背后的奥秘是科学研究人员一直孜孜以求的梦想。当人们认识到从基因到蛋白质到最后的生理功能这条信息链以后，生命科学的研究兴趣就更多地投向基因表达和生物性状之间复杂的联系上，基因组计划只是这一努力的第一步，功能性基因的发现和研究才是进一步的目标。差减杂交法提供了一种揭示基因和表型关系的手段，和其他分离基因的方法一样，差减杂交法经历了一个不断改进和完善的过程，这是一个良好的起点，但是就像人类的认识一样永远不会有终点。

（潘德思）

第三章

mRNA 差异显示技术

内容提要

mRNA DD－PCR 技术是筛选和克隆差异表达基因的一条快速而有效的途径。RNA 的质量、RT－PCR 的质量控制是该技术的关键。本章介绍了 mRNA DD－PCR 的基本原理和具体操作步骤。为了获得真正有意义的差异表达基因，对筛选出来的差异表达基因应进行杂交证实和序列测定，并利用某些数据库进行必要的分析。

mRNA 差异显示技术（mRNA differential display PCR，mRNA DD－PCR）是由 Peng Liang 等人在 1992 年建立的筛选基因差异表达的有效方法。它是将 mRNA 逆转录技术和 PCR 技术相结合的一种 RNA 指纹图谱技术。每一种组织细胞（包括同一组织细胞经过不同的处理）都有其特异表达的不同于其他组织细胞的基因谱，即特异的 RNA 指纹图谱。其基本原理是将基因背景相同的 2 个或几个细胞系或组织 mRNA 逆转录成 cDNA，用不同引物对，进行 PCR 扩增，扩增时加入同位素标记的核苷酸。用测序胶电泳分离 PCR 产物，经放射自显影即可找到被扩增的差异表达的基因。该方法具有快速、灵敏和重复性好等优点。

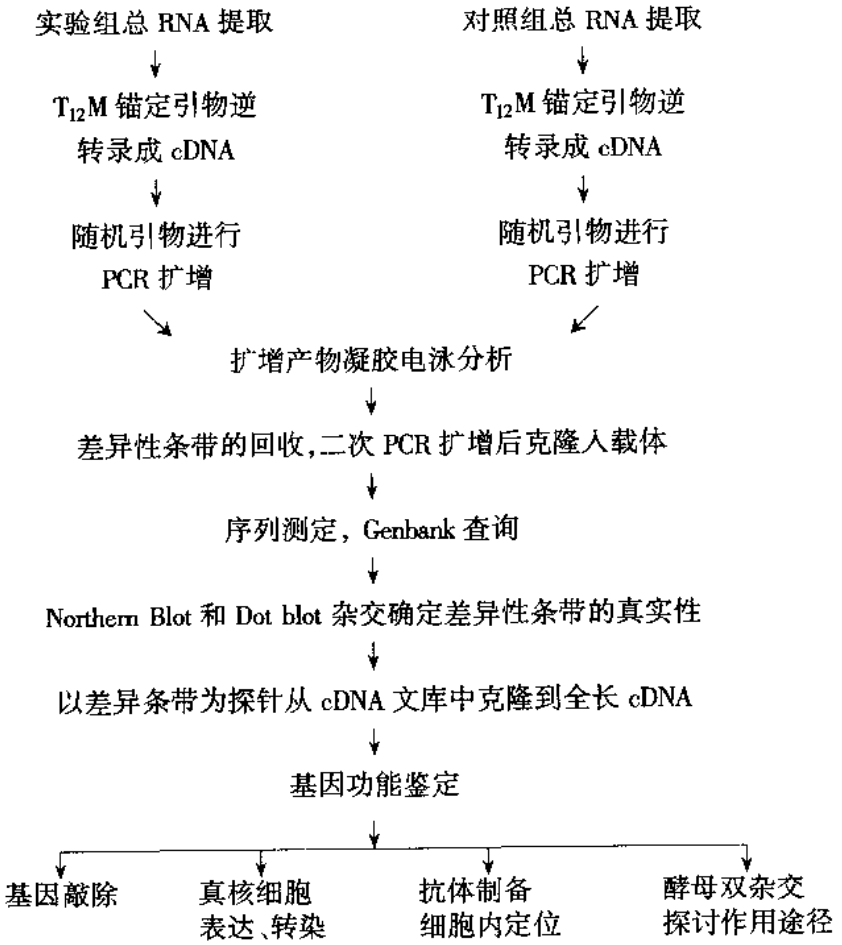
第一节　mRNA DD－PCR 的基本原理

差异基因表达是细胞分化的基础。正是这些基因在细胞中的特异表达与否，决定了生命历程中细胞的发育和分化、细胞周期的调节、细胞衰老和死亡等。mRNA DD－PCR 技术正是对组织特异性表达基因进行分离的一种快速而行之有效的方法。不同的组织/细胞或遗传背景相同的同一组织/细胞经过不同的处理后，提取各自的总 RNA，进行 mRNA 的逆转录合成 cDNA。进行逆转录时 3′端引物采用 oligo$(dT)_{12}$MN，其中 M 为 A、C、G 中的任何一种，N 为 A、C、G 或 T 中的任何一种，所以共有 12 种 oligo$(dT)_{12}$MN 引物。其中 M 称为锚定引物，起增大引物 Tm 值的作用。N 称为分类碱基，对逆转录进行归类。用这 12 种引物分别对同一总 RNA 样品进行 cDNA 合成，即进行 12 次不同的逆转录反应，从而使逆转录的 cDNA 具有 12 种类型，也就是对 cDNA 进行 12 种归类（目前较为流行的是进行 4 种归类，即 3′端引物采用 oligo$(dT)_{12}$M，M 为 A、C、G 中的任何一种）。5′端引物采用 20 条由 10 个碱基组成的随机引物。对每一类 cDNA 进行随机引物和逆转录引物 PCR 扩增，通过对不同的组织/细胞或经不同处理的同一组织/细胞的同一类 cDNA 的 PCR 选择性扩增产物的凝胶电泳分析，从而反映出不同样品的基因在

时间和空间上的组织特异性表达。对 cDNA 归类的目的是要降低随后所要进行的 PCR 扩增模板种类的多样性。因为通常情况下，在某一细胞中或某一个关键的发育时间点上有 15 000 种以上的基因在表达。如此众多种类的 mRNA 逆转录产物经 PCR 选择性扩增后其类型依然众多，很难用电泳系统加以快速准确地分离。因而需要对逆转录产物的 cDNA 进行归类处理，减轻不同 PCR 产物电泳分离的难度，提高分离的准确率。

不同的组织/细胞或经不同处理的同一组织/细胞的同一类 cDNA，经 PCR 选择性扩增的产物在凝胶电泳上的差别，表现在同一迁移位置上条带的有无和条带所含分子数目的多少。条带的有无反映出组织特异性表达的基因"开"与"关"。因为 mRNA 差异显示属半定量分析，所以条带中分子数目的多少，也是 mRNA 剂量效应的一种体现。

mRNA 差异显示技术的实验程序如下：



第二节　mRNA DD－PCR 技术的具体操作步骤

一、组织样品总 RNA 的提取

建议使用公司生产的总 RNA 提取试剂盒提取总 RNA。

称取 1 g 组织在液氮中研磨成粉

↓

将 12 mL 预冷的变性液倒入研钵中

↓

待液化后转入 50 mL 离心管中

↓

加入 1.2 mL 2 M NaAc (pH 4.0)
彻底混匀,上下颠倒 50 mL 离心管 4 ~ 5 次

↓

加入 12 mL 酚:氯仿:异戊醇
剧烈振荡 10 s,冰浴 20 min

↓

4 ℃离心,12 000 r/min × 20 min
上清转至另一 50 mL 离心管

↓

加入等体积异丙醇,混匀后置 -20 ℃ 30min 以上

↓

4 ℃离心,12 000 r/min × 10min
弃上清,加入预冷的 75% 乙醇 10 mL 洗涤沉淀

↓

4 ℃离心,12 000 r/min × 10 min
弃上清,加适量 RNase - free 水溶解沉淀

二、去除 RNA 样品中的微量 DNA

取 30 μg 总 RNA。应用 DNase I 去除其中残留的微量 DNA。反应体系如下:

10 × DNase I buffer	6 μL
Total RNA	30 μg
RNasin[40 U/ μL]	20 μL
DNase I (RNase - free) [1 U/ μL]	10 μL
RNase - free 水加至 60 μL	

↓

混匀后,室温静置 15 min

↓

加入 6.0 μL EDTA (25 mM,pH 8.0)

↓

65 ℃加热 10 min

↓

加入 45 μL 酚和 15 μL 氯仿:异戊醇(24:1)

↓

混匀,置冰浴 10 min

↓

4 ℃离心，12 000 r/min × 10 min

↓

回收上清，加入 6 μL 3M NaAc（pH 4.0）
和 120 μL 无水乙醇，混匀

↓

置 - 80 ℃ 45 min

↓

4 ℃离心，12 000 r/min × 15 min

↓

75% 乙醇洗涤沉淀

↓

稍干燥后加适量 RNase - free 水溶解
- 70 ℃保存备用

三、引物序列

3′端锚定引物：$Oligo(dT)_{12}A$：5′ - TTT TTT TTT TTT A - 3′
$Oligo(dT)_{12}G$：5′ - TTT TTT TTT TTT G - 3′
$Oligo(dT)_{12}C$：5′ - TTT TTT TTT TTT C - 3′

5′端随机引物：

AP - 1：	GTTTCGCTCC	AP - 2：	TGATCCCTGG
AP - 3：	CATCCCCCTG	AP - 4：	GGACTGGAGT
AP - 5：	TGCGCCCTTC	AP - 6：	TGCTCTGCCC
AP - 7：	GGTGACGCAG	AP - 8：	GTCCACACGG
AP - 9：	TGGGGGACTC	AP - 10：	CTGCTGGGAC
AP - 11：	CAGGCCCTTC	AP - 12：	TGCCGAGCTG
AP - 13：	AGTCAGCCAC	AP - 14：	AATCGGGCTG
AP - 15：	AGGGGTCTTG	AP - 16：	GGTCCCTGAC
AP - 17：	GAAACGGGTG	AP - 18：	GTGACGTAGG
AP - 19：	GGGTAACGCC	AP - 20：	GTGATCGCAG

四、逆转录归类反应体系的建立

分别取 1 μg 总 RNA，应用锚定引物 $T_{12}M$（M = G，A 或 C）将其逆转录成 cDNA。反应体系如下：

总 RNA	1 μg
$T_{12}M$ primer[10μm]	2 μL
RNasin[40 U/ μL]	1 μL
RNase - free water	12 μL

↓

70 ℃水浴，10 min

↓

置冰浴 1 min,短暂离心 10 s

↓

加入:5 ×第一链缓冲液	4 μL(试剂盒中含有)
0.1M DTT	2 μL
10 mM 脱氧核苷酸混合物	1 μL
SuperScript™ Ⅱ逆转录酶 [200 U/ μL]	1 μL

↓

42 ℃水浴,50 min

↓

70 ℃水浴,15 min

↓

置冰浴,加入 1 μL[2U] *E. Coli* RNase H

↓

37 ℃水浴,20 min

↓

保存于 -20 ℃备用

五、PCR 选择性扩增

各取经过某一锚定引物逆转录的实验组和对照组的 cDNA 2 μL,利用某一随机引物进行 PCR 扩增,反应条件如下:

10 × PCR buffer	2 μL
250 μM 脱氧核苷酸混合物	4 μL
5 μM 随机引物	2 μL
cDNA	2 μL
[α-^{32}P]dATP[10 μCi/ μL]	1 μL
Taq DNA 聚合酶[5 U/ μL]	1 μL
灭菌水加至	20 μL

↓

混匀,加 25 μL 矿物油

↓

94 ℃,5 min

↓

94 ℃, 30 s
40 ℃, 2 min
72 ℃, 30 s
} 进行 40 个循环

↓

72 ℃, 5 min

↓

保存于 4 ℃备用

六、mRNA DD－PCR 产物的变性聚丙烯酰胺凝胶电泳

取 3.5 μL PCR 产物加 2 μL 2× 上样缓冲液，80 ℃变性 2 min，置冰浴，每个样品上样于 6% 变性聚丙烯酰胺凝胶，恒定功率 60 W，电泳约 3 h，直至指示剂二甲苯青染料接近凝胶底部。揭下经硅烷化的一层玻璃板，在胶面平整地铺一层保鲜膜，然后在暗室内压一张 X 光片，X 光片的右边和底边与支撑凝胶的玻璃板的右边和底边对齐，以确定 X 光片与凝胶的相对位置。于 －70 ℃曝光 12～14 h 后洗片。如图 3－1 中箭头所示为差异 cDNA 条带。

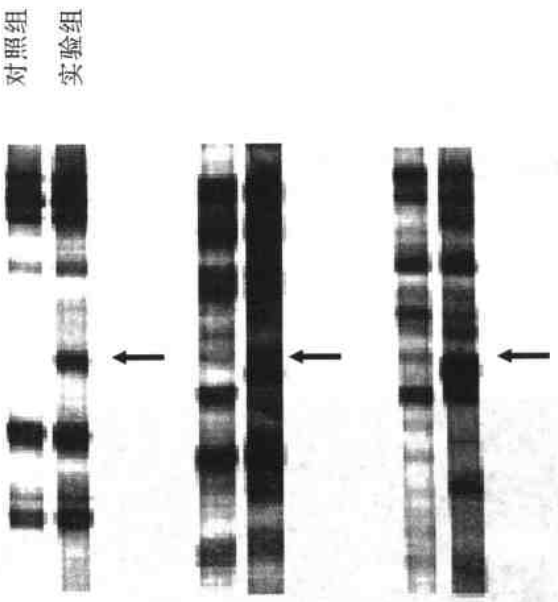


图 3－1　mRNA DD－PCR 产物变性聚丙烯酰胺凝胶电泳结果
（摘自 Yang DC, et al. Cir Res, 1999, 84:704～712）

七、差异表达条带的回收

对 mRNA DD－PCR 产物放射自显影图谱进行比较分析。将差异表达的 cDNA 条带从聚丙烯酰胺凝胶上切下来，放入 1.5 mL 离心管中，然后按下述步骤操作：

用 1 mL TE buffer 快速冲洗一遍，然后加入 100 μL TE buffer

碾碎胶条，室温放置 15 min 后，100 ℃水浴 15 min

↓

4 ℃离心，12 000 r/min × 2 min

↓

上清转入另一 1.5 mL 离心管中

↓

加入：10 μL 3M NaAc（pH 4.8），5 μL 糖原（10 mg/mL）
450 μL 无水乙醇
↓
－20 ℃沉淀过夜
↓
4 ℃离心，12 000 r/min × 15 min
↓
弃上清，沉淀加入预冷的 85% 乙醇 200 μL
↓
离心 5 min 后去上清
↓
沉淀稍干燥后加入 10 μL 双蒸水溶解，－20 ℃保存备用

八、回收条带的再次扩增

取 4 μL 回收的 cDNA，用同一随机引物进行二次 PCR 扩增。图 3－2 所示为二次 PCR 扩增结果。

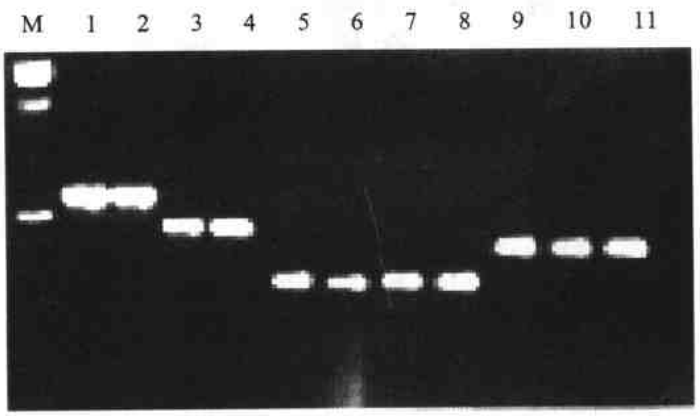


图 3－2　差异 cDNA 条带二次 PCR 扩增结果
（摘自 Yang DC, et al. Cir Res, 1999, 84:704－712）

二次 PCR 扩增及其扩增产物回收的操作步骤为：

10 × PCR 缓冲液	5 μL
250 μM 脱氧核苷酸混合物	25 μL
5 μM 随机引物	1 μL
cDNA	4 μL
Taq DNA 聚合酶[5 U/ μL]	1 μL
灭菌水加至	50 μL
矿物油	50 μL

↓

第一次 PCR 扩增：94 ℃，10 min

↓

94 ℃，1 min
40 ℃，1.5 min
72 ℃，1.5 min

↓ 进行 20 个循环

72 ℃，10 min

↓

第二次 PCR 扩增：94 ℃，10 min

↓

94 ℃，1 min
42 ℃，1.5 min
72 ℃，1.5 min

↓ 再进行 20 个循环

72 ℃，10 min

↓

取 15 μL 于 2% 琼脂糖电泳检测

↓

在剩余的 PCR 产物中加双蒸水至 100 μL

↓

加入 100 μL 酚∶氯仿∶异戊醇(25∶24∶1 V/V)

↓

4 ℃离心，12 000 r/min × 10 min，取上清

↓

加入：10 μL 3 M NaAc (pH 5.2)
5 μL 糖原(10 mg/mL)
250 μL 无水乙醇

↓

-20 ℃过夜

↓

4 ℃离心，12 000 r/min × 15 min

↓

弃上清，沉淀用 75% 乙醇

↓

室温离心，12 000 r/min × 10 min

↓

沉淀再加 75% 乙醇

↓

再次室温离心，12 000 r/min × 10 min

↓

将沉淀溶于适量无菌双蒸水中，-20 ℃保存备用

九、二次扩增产物的克隆、测序和 GenBank 查询

采用 Promega 公司的 pGEM-T Easy Vector System I 试剂盒，按试剂盒说明书将二次 PCR 扩增产物克隆至 pGEM-T Easy 载体上(图 3-3)。宿主菌为 JM109。经转化铺盘后，在培养盘上挑选白色菌落，摇菌后提取质粒，经 *EcoR* I 酶切鉴定。

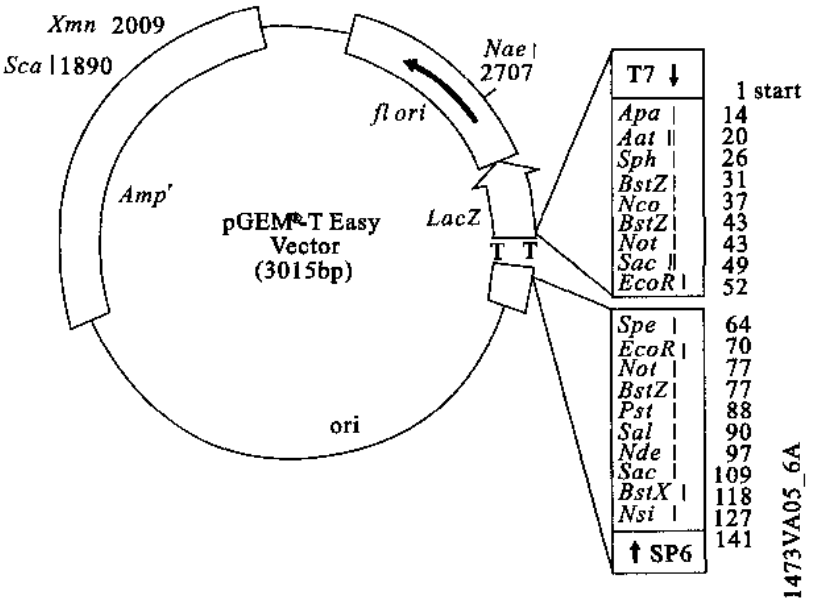


图 3-3　pGEM-T Easy 载体的结构

采用 ABI377 自动测序仪，应用 T7 或 SP6 通用引物，对克隆的片段进行序列测定。

应用测序所得的 cDNA 序列，在 NCBI 的 GenBank 数据库中进行 BLAST 查询，以了解该序列有关的信息。如该序列是否为新基因？数据库中是否存在其相应的或同源的 EST 或全长 cDNA？其相应或同源基因的功能是什么？这样就为进一步的克隆和功能研究提供了参考。

十、Northern Blot 和 Dot Blot 杂交

采用 Northern Blot 和 Dot Blot 杂交检验和证实所得的差异条带 cDNA 确实存在差异。其中 Northern Blot 杂交还可预测差异条带 cDNA 的大小，为将来筛选和克隆全长 cDNA 提供参考。

(一)Northern Blot 杂交

1. 准备杂交膜。

(1)配制 1%琼脂糖甲醛变性胶。

10×MOPS 缓冲液	10 mL
DEPC 处理水	90 mL

琼脂糖　　1 g

加热至胶完全溶化后，置于室温冷却至 60 ℃，依次加入 6 μL 溴乙锭 EB(10 mg/mL)，5.4 mL 37% 甲醛，混匀并置室温 15 min 后制胶。

(2)电泳。20 μg 的实验组和对照组总 RNA 样品加适量上样缓冲液，95 ℃水浴变性 2 min，置于冰上。同时将胶放入装有 1 × MOPS 缓冲液的电泳槽中预电泳 10 min。然后加样，在 100 V 电压下电泳，直到溴酚蓝染料迁移至胶下游 3/4 处。电泳结束后，将胶置于紫外灯下同荧光尺一起拍照，以记录 28S rRNA 和 18S rRNA 的位置，作为 RNA 分子量的参照。真核细胞 28S rRNA 和 18S rRNA 的分子量分别约为 5 kb 和 2 kb。

(3)RNA 的转移和固定。

1)将胶放入 0.05mmol NaOH 中浸泡 20 min。

2)用 DEPC 处理水冲洗胶一次，再用 20 × SSC 浸泡 45 min，以除去胶中甲醛。

3)用毛细管转移法将胶上 RNA 吸引至尼龙膜，转移液为 20 × SSC，转移时间至少 24 h。

4)转移结束后，用 5 × SSC 漂洗尼龙膜数分钟，取出尼龙膜，置于滤纸上，室温晾干，紫外交联 2 min，然后在 80 ℃烤箱真空干烤 2 h。

5)将烤干的尼龙膜封于塑料袋中，存于 - 20 ℃备用。

2.探针的制备和标记。

(1)探针的制备。将克隆到 pGEM - T Easy 载体中的差异条带 cDNA 用 *EcoR* Ⅰ或 *Not* Ⅰ切割下来，然后在 1% 琼脂糖凝胶电泳上，用 Bio Rad 公司生产的 DNA 回收试剂盒，按操作说明书回收差异条带 cDNA，最后用 20 μL 灭菌水洗脱，保存在 - 20 ℃用于以后同位素标记。

(2)标记探针。应用 Promega 公司生产的随机引物探针标记试剂盒进行标记。其标记过程如下：

cDNA (约 25 ng) + H_2O → 总体积 9 μL

↓

100 ℃水浴变性 10 min

↓

置于冰上冷却

在上述管中依次加入下列成分：

dATP, dGTP, dTTP 混合物	3 μL
10 × 反应缓冲液	2 μL
[α - ^{32}P] - dCTP[10μCi/ μL]	5 μL
Klenow 酶[2 U/ μL]	1 μL

总体积 20 μL

↓

37 ℃水浴 30 min

↓

65 ℃水浴 10 min，以灭活酶

↓

保存于 - 20 ℃备用

3.杂交。

(1)杂交液的配制。

1)准备 5×"P"缓冲液:

1.0%牛血清白蛋白、1.0% 聚乙烯吡咯烷酮(PVP,M.W.40 000)。

1.0% 聚蔗糖(ficoll,M.W.400 000)、250mM Tris - HCl (pH 7.5)、0.5% 焦磷酸钠和 5.0% SDS。

2)准备 50%(w/v)的硫酸葡聚糖

3)准备 10 mL 的杂交液:

去离子甲酰胺	5 mL
5×"P"缓冲液	2 mL
50% 硫酸葡聚糖	2 mL

总体积 9 mL

↓

颠倒混合

↓

42 ℃水浴 10 min

↓

加入 0.58 g NaCl, 混合

↓

42 ℃水浴,直到 NaCl,完全溶解

↓

加入 1 mL 经 100 ℃水浴煮沸 10 min,
然后在冰上冷却的鲑精 DNA (1 mg/mL)

↓

混合后即可用于预杂交和杂交

(2)杂交的操作过程。

1)预杂交。在封有尼龙膜的塑料袋一角剪一口,加入杂交液 10 mL,然后封上口,42 ℃水浴摇床过夜。

2)杂交。将标记好的探针 100 ℃水浴煮沸 10 min,然后置于冰上冷却。在含有尼龙膜的预杂交塑料袋一角剪一口,倒出预杂交液。将探针和新配制的杂交液 10 mL 混合后加入塑料袋中,封上口。42 ℃水浴摇床过夜。

(3)洗膜。从杂交袋中取出尼龙膜,按下列步骤洗膜:

250 mL 2×SSC 室温下洗膜二次,每次 10 min

↓

500 mL(2×SSC + 1.0% SDS)于 65 ℃水浴洗膜二次,每次 30 min

↓

250 mL 0.1×SSC 室温下洗膜二次,每次 15 min

↓

将膜用滤纸吸干,封于塑料袋中

(4)放射自显影。用 X 光片放射自显影,-70 ℃曝光 3~4 d。

图 3-4 所示为 Northern Blot 杂交结果。箭头所指为在实验组中高表达而在对照组中无表达的差异 cDNA。

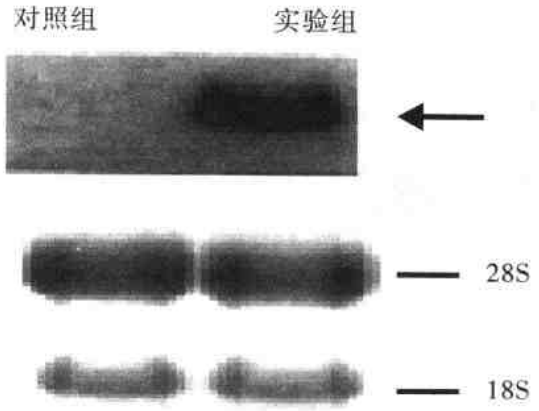


图 3－4　Northern Blot 杂交结果
（摘自 Yang DC, et al. Cir Res, 1999, 84:704～712）

（二）Dot Blot 杂交

提取含有克隆片段的质粒 DNA。取 5 ng 分别加入 NaOH 和 EDTA，使其终浓度分别为 0.4 M和 10 mM。100 ℃加热 10 min 使其充分变性，置冰浴，然后加入等体积预冷的 2 M NaAc（pH 7.0）。将变性后样品点样于尼龙膜上，紫外交联 10 min，将尼龙膜放入适当大小的杂交盒中，加入适量预杂交液（5 × Denhardt's，0.5% SDS，6 × SSC，50% 甲酰胺，100 μg/mL 变性鲑精 DNA）。42 ℃水浴摇床预杂交过夜。弃预杂交液，加入等体积杂交液（成分与预杂交液相同），在两个杂交盒中分别加入经过变性的以随机引物法标记的实验组和对照组的总 cDNA，42 ℃水浴摇床杂交 24 h。洗膜：2 × SSC，0.5% SDS，室温洗膜 5 ～ 10 min；2 × SSC，0.1% SDS，室温洗膜 10 ～ 15 min；0.1 × SSC，0.5% SDS，42 ℃洗膜 30 min；0.1 × SSC 室温短暂漂洗滤膜，置于一叠纸巾上，吸去液体，压片进行放射自显影，曝光时间为 48 h 左右。图 3－5 所示为 Dot Blot 杂交结果，箭头所指为在实验组中高表达而在对照组中无表达或低表达的差异 cDNA。

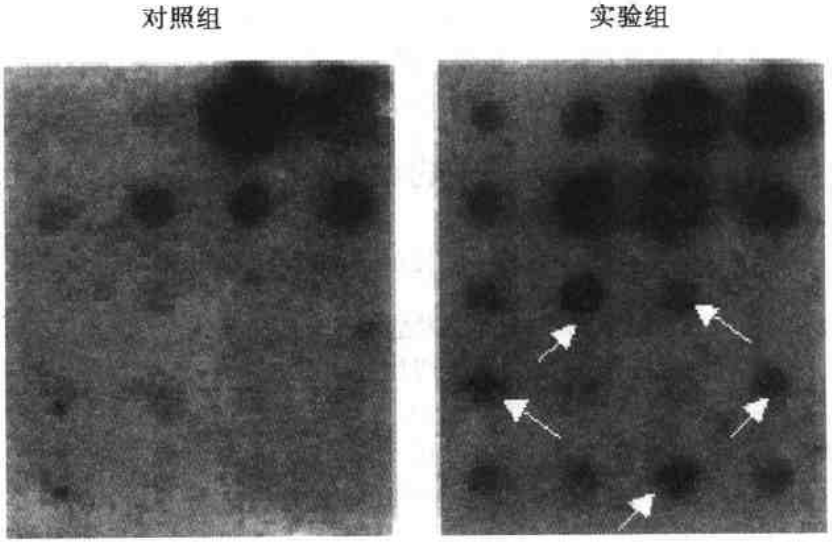


图 3－5　Dot Blot 杂交结果
（摘自 Yang DC, et al. Cir Res, 1999, 84:704～712）　（魏英杰）

第四章

荧光原位杂交(FISH)技术

内容提要

荧光原位杂交(FISH)是一种新的非放射性原位杂交技术。具有安全、灵敏、快速和分辨率高等优点。FISH技术利用非放射性标记核酸代替放射性标记核酸在染色体、细胞或组织切片上进行原位杂交,通过免疫细胞化学的过程以荧光信号对杂交体加以检测和定位。近年来FISH技术在分辨率、灵敏度、容量及特异性方面有进一步的发展,它不但用于基因的染色体定位和高分辨物理图谱的绘制,而且是研究分裂期和间期细胞染色体结构和数量变化的有效方法,是分子生物学与细胞遗传学的桥梁技术。

发展染色体基因定位技术不仅有助于克隆疾病相关基因,也有助于人们认识染色体的结构及研究基因的功能。该技术的发展经历了几个阶段。首先体细胞杂交技术的建立可将特定的序列定位在某条染色体上,如果用高能射线处理杂交细胞,使染色体断裂成若干片段,还可以利用遗传标记将目标基因定位在染色体的某一区段内。1974年Evans首次将染色体显带技术和染色体原位杂交联合应用,提高了定位的准确性。20世纪70年代后期人们开始探讨荧光标记的原位杂交,即FISH技术。1981年Harper成功地将单拷贝的DNA序列定位到G显带标本上,标志着染色体定位技术取得了重要进展。20世纪90年代,随着人类基因组计划的进行,由于绘制高分辨人类基因组图谱的需要,FISH技术得到了迅速的发展和广泛应用。

第一节　FISH的原理和技术特点

一、FISH技术的原理和优点

FISH (fluorescence *in situ* hybridization)技术是一种重要的非放射性原位杂交技术。它的基本原理是:如果被检测的染色体或DNA纤维切片上的靶DNA与所用的核酸探针是同源互补的,二者经变性-退火-复性,即可形成靶DNA与核酸探针的杂交体。将核酸探针的某一种核苷酸标记上报告分子如生物素、地高辛,可利用该报告分子与荧光素标记的特异亲和素之间的免疫化学反应,经荧光检测体系在镜下对待测DNA进行定性、定量或相对定位分析。

原位杂交的探针按标记分子类型分为放射性标记和非放射性标记。用同位素标记的放射性探针优势在于对制备样品的要求不高,可以通过延长曝光时间加强信号强度,故较灵敏。缺点是探针不稳定、自显影时间长、放射线的散射使得空间分辨率不高、及同位素操作较繁琐

等。采用荧光标记系统则可克服这些不足,这就是 FISH 技术。该技术不但可用于已知基因或序列的染色体定位,而且也可用于未克隆基因或遗传标记及染色体畸变的研究。在基因定性、定量、整合、表达等方面的研究中颇具优势。

FISH 技术作为非放射性检测体系,具有以下优点:① 荧光试剂和探针经济、安全;② 探针稳定,一次标记后可在两年内使用;③ 实验周期短、能迅速得到结果、特异性好、定位准确;④ FISH 可定位长度在 1 kb 的 DNA 序列,其灵敏度与放射性探针相当;⑤ 多色 FISH 通过在同一个核中显示不同的颜色可同时检测多种序列;⑥ 既可以在玻片上显示中期染色体数量或结构的变化,也可以在悬液中显示间期染色体 DNA 的结构。

二、FISH 技术的基本操作过程

FISH 的基本操作过程与其他 DNA 杂交技术一样,即荧光标记的探针 DNA 与互补的染色体靶 DNA 结合,进行靶 DNA 的检测。

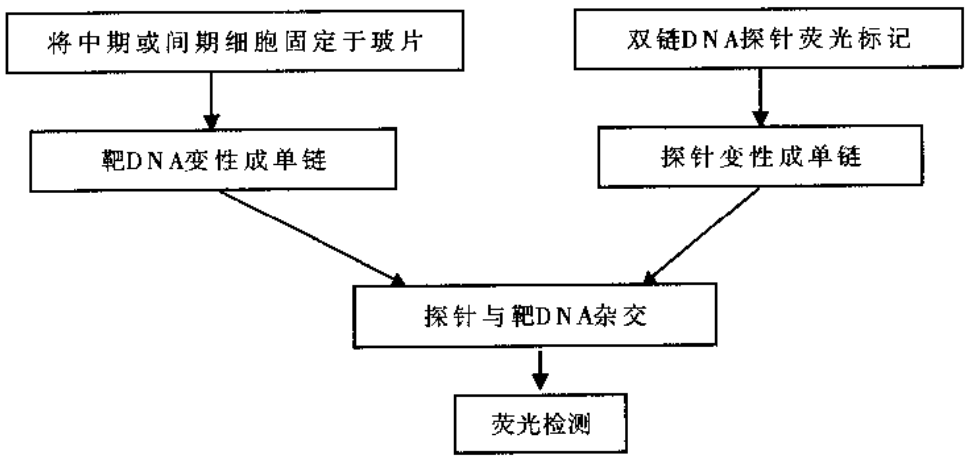


图 4-1　FISH 的操作流程

(一)FISH 标本的制备

FISH 的靶 DNA 固定于载玻片上的细胞核内。细胞可以是贴壁细胞或悬浮细胞,是分裂期细胞或间期细胞。也可以是血涂片、冰冻或包埋固定的组织切片。细胞用传统显带使用的方法低渗并固定于玻片上,将玻片在甲酰胺溶液中温育,使 DNA 解链,然后固定在冷乙醇中以防止在加入探针前 DNA 链退火。

(二)探针的制备

常用的探针按序列组成可分为 3 类:① 染色体特异的重复序列探针,如卫星 DNA。用于检测间期细胞非整倍体和微小标志染色体;② 全染色体或染色体区域特异性探针,来自可插入大片段的载体文库或微切割 DNA 的 PCR 产物。如 Alu-PCR,可用于中期染色体重组和间期核结构分析;③ 特异位置序列探针,常由一个或几个克隆序列组成,进行克隆 DNA 的染色体定位及检测靶 DNA 的序列拷贝数及结构变化。

由于一些重复序列可以作为染色体定位较高分辨率的标志,在进行基因定位或染色体畸变研究时,应当制备一个探针库,该探针库应尽量选取在整个人类染色体上能产生比较均匀的

杂交信号的分散重复序列，如 Alu 序列。探针库与特定的细胞染色体杂交将产生一个相对比较固定的杂交模式，而比较在不同条件下杂交模式的改变，就可发现需要的信息。有 3 种方式制备探针库：① Alu－PCR，Alu 序列为人类所特有，可避免异源 DNA 的干扰；② 简并寡聚核苷酸引物 PCR，又称 DOP－PCR（degenerate oligonucleotide primed－PCR）；③ 非序列依赖性扩增（sequence independent amplification，SLA）。后两者的扩增产物无种属特异性。

FISH 信号检出率与 DNA 标记探针长度有关。Yokota H 等研究发现，来自 cDNA 文库等短基因组序列（1～5 kb）的探针杂交信号检出率仅为 5%～20%。来自粘粒（cosmid）噬菌体克隆其片段（10～40 kb）检出率在 80%以上。而采用 YAC 文库的基因组 DNA（100～500 kb），其检出率达 95%。因此大片段的基因组 DNA 最适合做 FISH 的探针。为了抑制其中的重复序列带来的非特异杂交背景，可用未标记的基因组 DNA 或经 *Cot* I 酶切的基因组 DNA 先进行预杂交。

（三）探针标记

标记探针有两种基本方式：①直接标记：将荧光分子直接标记于探针 DNA/RNA 上。荧光分子可共价结合于核苷或磷酸戊糖骨架上，将荧光标记的核苷三磷酸采用缺口平移法掺入探针。这种方式快速、简便，背景干扰少，但杂交信号弱，因而不够灵敏。由于该标记方法所标记的探针种数不受高亲和力配基能力的限制，因此在多色 FISH 中应用更广。②间接标记法：采用某一中间分子标记探针，杂交后再用与该中间分子的亲和物或抗体耦联的荧光分子进行检测，这样可使被检信号放大，能检测到小至 500 bp 的靶序列。常用的中间分子有：生物素（biotin）、地高辛配基（digoxingenin）以及二硝基苯酚（dinitrophenyl）。常用的荧光标记分子及性质见表 4－1。

表 4－1　常用于标记的荧光分子及其性质

荧光分子	最大激发光（nm）	最大发射光（nm）	荧光
AMCA	358	450	蓝
异硫氰荧光素（FITC）	495	515	绿
罗丹明（rhodamin）	550	575	红
德州红（Texas－red）	595	615	深
Cy－5（红）	650	670	不可见红外光

掺入探针的方法通常采用缺口平移法或随机引物法。在探针 DNA/RNA 序列已知的情况下也可采用 PCR 或逆转录法。对于小于 1 kb 的探针，用 PCR 方法容易标记成功。

（四）杂交

标记的探针变性后，在甲酰胺溶液中与标本 DNA 杂交。最适的杂交条件主要取决于探针与靶基因的性质。对于染色体重复序列的检测，可直接将探针混合液置于玻片上，37 ℃孵育 15 min。对于单拷贝基因序列的检测，一般在杂交之前进行预复性，即与过量的基因组 DNA 或 *Cot* I DNA 先预杂交。由于重复序列与重复序列的杂交要比单拷贝序列与重复序列的杂交速度快，因此探针中非特异的重复序列可被封闭，以减少非特异杂交，然后再与靶 DNA 杂交

过夜。杂交结束后，标本经清洗去掉错配或未杂交的探针，然后用荧光标记的检测试剂进行检测。

(五)染色体显带

为了能客观清楚地辨认杂交位点，同时在研究中采用显带与 FISH 是十分必要的。常用的显带方法有：G 带、R 带、Q 带等。用氮芥奎吖因(Quinacrine mustard)产生的 Q 带，经过一定的预处理，如用放线菌素/DAPI 处理，再以 Giemsa 染色可显示 C 带；染色体经溴尿嘧啶处理，Hoechest 染色可显示 R 带；采用 Alu 或 LINE 重复序列与探针同时杂交，可呈现 R 或 G 带。这样就能辨认探针在染色体上的杂交部位。但杂交前显带将会大大降低杂交信号，原因是原位杂交与染色体显带存在操作上的矛盾，即染色体变性与显带相互影响。目前尚无理想的解决方法。

(六)荧光显微镜检测

显微镜的质量及滤片的选择对获得满意的观测结果至关重要，特别是要注意选择合适的激发/阻挡滤片组合。用多色滤光片仅一次曝光即可拍摄镜下所见，简化了观察和照相过程，但灵敏度不如单色滤光片高。数字成像显微镜由于可以分析信号荧光的光谱组成而将不同的荧光区分开，敏感性高，而且可进行量化分析和自动化分析，因此越来越多的被应用于 FISH 信号的检测。应用共聚焦显微镜可通过对染色体标本的不同平面进行断层扫描而获得高质量的图像，空间定位效果好，但目前它的灵敏度不如普通荧光显微镜。

(七)结果分析

尽管 FISH 的杂交效率很高，但也不能达到 100%杂交，特别是在应用较短的 cDNA 探针时效率明显下降。样品制备时由于各细胞染色质被压缩的程度不完全一致，使得细胞与细胞间信号的定位有轻微差异，特别是中期细胞，因此需要多个细胞才能将探针定位到染色体亚带上。对单拷贝 cDNA 或 DNA 序列的定位需要的精确度最高，需要综合细胞遗传学、连锁分析和物理图谱的结果。在基因定位的结果报告中应包括以下内容：① 杂交信号阳性细胞的比例；② 位点特异信号的比例；③ 背景信号的位置；④ 基因或 DNA 序列的精确定位。在进行细胞遗传学分析时，对染色体带型的判断具有一定的主观性，因此最好有两位以上的研究者同时进行观察。

第二节　FISH 技术的实验操作步骤

一、染色体标本的制备

(一)制备中期染色体标本

1. 淋巴细胞培养。取受检者外周血 0.5 mL 注入 5 mL 含有 20%小牛血清的 RPMI－1640 培养液体中，加 0.25 mg/mL 的 PHA－M 使用液 0.2 mL(终浓度为 10 μg/mL)，轻轻摇匀，置 37 ℃恒温箱中培养 72 h。

2. 终止培养。加 4 μg/mL 的秋水仙酰胺使用液 50 μL(终浓度为 0.04 μg/mL)，轻轻摇匀，继续培养 60 min，使细胞有丝分裂中止在分裂中期。

3. 收获细胞。室温离心，1 000 r/min×5 min，弃去上清。

4. 低渗。用预温至 37 ℃的 0.075 M 的氯化钾 8 mL 将沉淀轻轻吹打成细胞悬液，在 37 ℃

水浴中低渗 20 min。加新鲜配制的甲醇/冰醋酸(3:1)固定液 1 mL 预固定，室温离心，1 000 r/min×5 min，弃去上清。

5. 固定。用新鲜配制的甲醇/冰醋酸(3:1)固定液 8 mL 将沉淀轻轻吹打成细胞悬液，室温下固定 10 min，室温离心，1 000 r/min×5 min，弃去上清，共 2 次。

6. 滴片。取 0.5 mL 固定液将沉淀轻轻吹打成细胞悬液，于 0.3 m 高处滴 3～4 滴于预先保存在冰水中的冷湿载玻片上，并立即吹气使染色体分散，在酒精灯上过火 2～3 次，气干。

注意：制备良好的中期染色体分裂相是中期 FISH 成功的基本保证。秋水仙素的作用时间也很重要，一般 1 h 为易，作用时间长，染色体过度收缩，导致染色体缩短，难以辨认；作用时间短，则分裂相太少。另外，若制片后染色体分相不好，可用蒸汽熏蒸 2～3 s，然后立即放 4 ℃冰箱，10 min 后转到 -70 ℃保存。

(二)制备高分辨的染色体标本(5-氟尿嘧啶核苷同步化法)

1. 细胞培养。外周血淋巴细胞常规培养 72 h(同上)。

2. 同步化。加 5-氟尿嘧啶核苷(5-Fdu)使用液 50 μL 及尿嘧啶使用液 50 μL(终浓度分别为 10 μM 和 4 μM)，摇匀，37 ℃继续培养 17 h。

3. 阻断。加入 Brdu 使用液 30 μL(终浓度为 30 μg/mL)，摇匀，37 ℃继续培养 7 h。

4. 终止培养。加 4 μg/mL 的秋水仙酰胺使用液 50 μL(终浓度为 0.04 μg/mL)，轻轻摇匀，继续培养 10 min。

5. 收获细胞、低渗、固定及滴片等操作同上。

(三)制备伸展的染色质标本

有数种伸展染色质的制备方法，包括碱抽提法、去污剂-高盐法、SDS 法 NaOH-乙醇法以及甲酰胺处理法等。它们对细胞的裂解程度不同，产生不同伸展程度的 DNA。下面以 NaOH-乙醇法为例予以介绍：

1. 按前述方法制备、收集丝裂原刺激的淋巴细胞，但不用秋水仙酰胺处理。

2. 将细胞悬液滴加到玻片上，在完全干燥前，将玻片浸入 PBS 中 1 min，用吸水纸吸去多余的水分，立即进行碱处理。

3. 碱处理。准备好干净的盖玻片，在玻片的一端加碱处理液 100 μL，将玻片倾斜 30 度，用盖玻片的一边放在碱处理液中，向玻片的另一端拉过，使碱处理液均匀的分布在玻片上，用吸水纸吸去多余的液体。(碱处理液：0.07 M NaOH 和无水乙醇按 5:2 混匀。)

4. 玻片平放，用少量甲醇小心淋洗数次。

5. 空气干燥，在 70%、80%、95%的梯度乙醇中，脱水各 2 min。

6. 空气干燥，储存于 -20 ℃的干燥箱内。

注意：甲酰胺处理法为使用 70%的甲酰胺 2×SSC 溶液，其余步骤与碱处理法相同。碱处理法可以完全破坏细胞核，使玻片表面的 DNA 分子呈网状。甲酰胺法要温和一些，适合较长探针的杂交。

二、染色体显带

此步操作通常用于常规染色体核型分析，以便与 FISH 结果进行对比研究。也可在显带以后继续进行 FISH 操作，但显带后会影响 FISH 的杂交信号。

1. 胰酶消化。将干燥的玻片在预温至 37 ℃的胰蛋白消化液中轻轻摆动 3～10 s，立即在

37 ℃生理盐水中洗 2 次。

2.染色。置 37 ℃ Giemsa 染液中染色 5 ~ 10 min,自来水冲洗,晾干。

3.镜检。在光学显微镜下,采用常规染色体 C 显带(320 ~ 550 条带)和高分辨 G 显带(550 ~ 800 条带)技术进行核型分析,每份标本分析 30 个中期或早中期分裂相。

三、Alu – PCR 法制备探针

1.PCR 模板:用标准方法制备的基因组 DNA 或其他 DNA 大片段。

引物:5′ GGATTACAGGYRTGAGCCA

3′ RCCAYTGCACTCCAGCCTG

(R 代表 A/T, Y 代表 C/G)

2.PCR 反应体系:

模板 DNA	100 ng
引物	1.00 mM (终浓度)
dNTP	0.25 mM(终浓度)
10 × PCR 缓冲液	5.0 μL
25 mM $MgCl_2$	4.0 μL
Taq 酶	2.5 U
补 ddH_2O 至	50.0 μL

3.PCR 反应条件:95 ℃变性 5 min,随后 95 ℃,1 min, 65 ℃ 1 min, 72 ℃ 3 min 共 30 个循环,然后 72 ℃ 延伸 10 min。

4.PCR 产物经 Micro100 柱纯化后标记。

四、切口平移法标记探针

1.主要试剂。

(1)反应缓冲液:0.5 M Tris – HCl,pH 7.5 ,0.1 M $MgCl_2$。

(2)dNTP 混合液:dATP, dGTP, dCTP 各 0.3 mM, 溶于 50 mM Tris – HCl pH 7.5。

(3)DNase I:6.5 mg/mL 溶于 0.1 M $MgCl_2$。

2.在 Eppendorf 管中加入下列试剂:

探针	1.0 μg
10 × 反应缓冲液	5.0 μL
10 × dNTP	5.0 μL
0.3 M Bio – 16 – dUTP	2.5 μL
(或 0.3M Dig – 11 – dUTP)	
DNA 聚合酶	20 U
DNase I	0.05 μL
加 ddH_2O 至	50.0 μL

3. 充分混匀，14 ℃水浴 2 h，将反应管置于冰浴中。

4. 取 5 μL 反应液进行琼脂糖凝胶电泳确定探针长度，理想的长度为 300 ~ 600 bp，若太长，补加适量 DNase I，14 ℃继续水浴 0.5 h；若太短，减少 DNase I 的用量重新标记。

5. 在反应管中加入 1 μL 0.5mmol/L EDTA，1 μL 5% SDS，65 ℃加热 10 min 以终止反应。

6. 用 Sephadex G - 50 柱去除未标记的核苷酸。

五、FISH 检 测

1. 制片。按常规方法制备中期染色体标本，用钻石刻刀在玻片上划出 25 mm × 25 mm 的拟杂交区域。

2. 玻片预处理。在 37 ℃ 2 × SSC 中温浴 30 min，室温下用 70%、85%、100% 系列酒精脱水各 2 min，气干。

3. 变性置 70 ℃ 70% 甲酰胺/2 × SSC 变性液中变性 2 min 后立即置冷藏至 - 20 ℃的 70%、85%、100% 乙醇中系列脱水各 2 min，空气中自然干燥。

4. 杂交将地高辛标记的探针 10 μL 置 37 ℃水浴 5 min，加于变性的染色体标本拟杂交区域，玻璃盖玻片覆盖，用家用胶或凝胶封片后置湿盒中 37 ℃作用 20 ~ 22 h。

注意：盖玻片盖上去后不能再移动，否则有可能破坏分裂相。封片一定要严，有条件的用 Rubber Cement 封片最好。

5. 漂洗。去除盖玻片，依次在 43 ℃ 50% 甲酰胺/2 × SSC 漂洗液中手摇漂洗 15 min，37 ℃ 2 × SSC 中手摇漂洗 8 min，室温下 1 × PBS 中漂洗 5 min × 2 次，稍干。

6. 荧光染色。加 FITC 标记的抗地高辛 60 μL，用塑料盖玻片覆盖，置 37 ℃湿盒中作用 5 min，1 × PBS 洗 2 min × 3 次。此时可用碘丙锭复染后在荧光显微镜下观察，如杂交信号太弱，可在加碘丙锭之前应用间接免疫荧光法进行信号扩增。本实验中均行信号扩增(如下述)。

7. 信号扩增。依次加兔抗羊抗体 60 μL，用塑料盖玻片覆盖，置 37 ℃湿盒中作用 15 min，1 × PBS 漂洗 3 min × 3 次。加 FITC 标记的抗兔抗体 60 μL，塑料盖玻片覆盖，再放入 37 ℃湿盒中作用 15 min，1 × PBS 漂洗 3 min × 3 次，气干。

注意：能否得到信号强、背景低的 FISH 结果，这是关键的一步。整个过程要保持标本湿润，否则会造成高背景。若为小片段标本则需 2 ~ 3 次放大，方可显示出杂交信号。

8. 复染。取碘丙锭(PI)及抗褪色剂各 6 μL，混匀后点在杂交区域的四角，盖 25 mm × 25 mm 玻璃盖玻片，轻压以排除气泡。

9. 镜检及照相。在 Nikon FX - 35DX 荧光显微镜下观察。激发波长 430 nm，阻滞波长 510 nm，计数杂交信号，每份标本计数 20 个分裂相，用 Kodak 400 负片拍照。

在荧光显微镜下可见染色体被染成均匀的橘红色，以 FITC 标记的杂交信号为明亮的黄绿色荧光。(图 4 - 2，图 4 - 3 显示的是 22q11 特异的 N25 DNA 探针对先天性心脏锥干畸形患者进行 22q11 微缺失的检测结果)。

杂交后标本应尽快进行检测，迅速寻找到好的杂交信号拍照，在荧光显微镜下，每一次激发都会减弱信号强度，不观察时应立即关闭显微镜。若杂交后不能立即观察，应将标本平放暗盒中，- 20 ℃保存不超过 2 周。

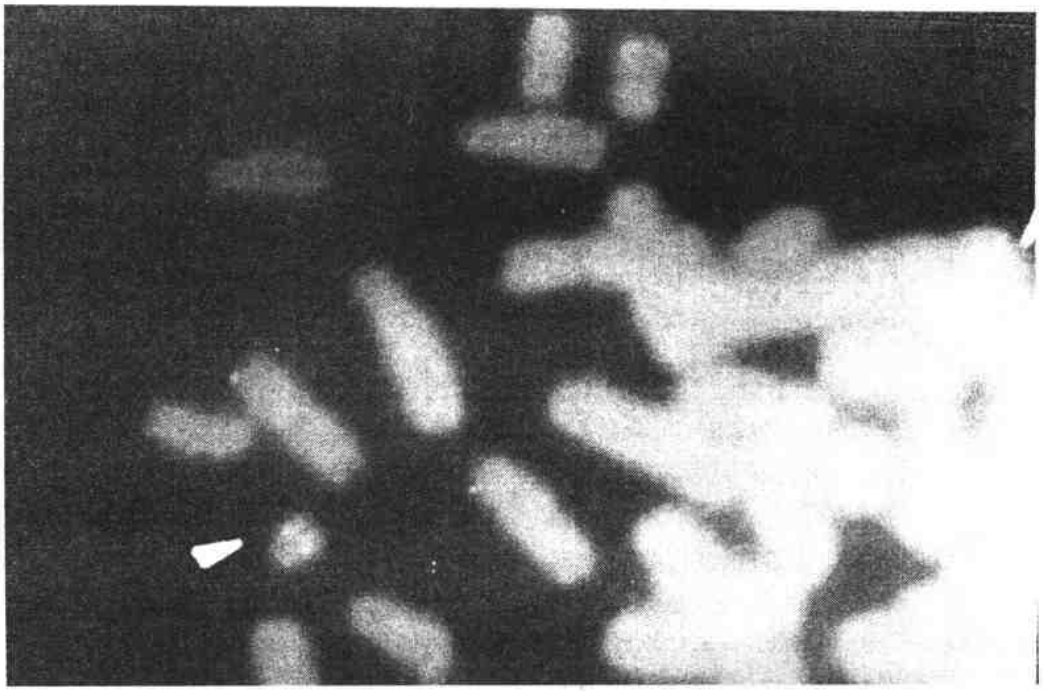


图 4－2　无 22q11 微缺失，2 条 22 号染色体均可见杂交信号(箭头所示)

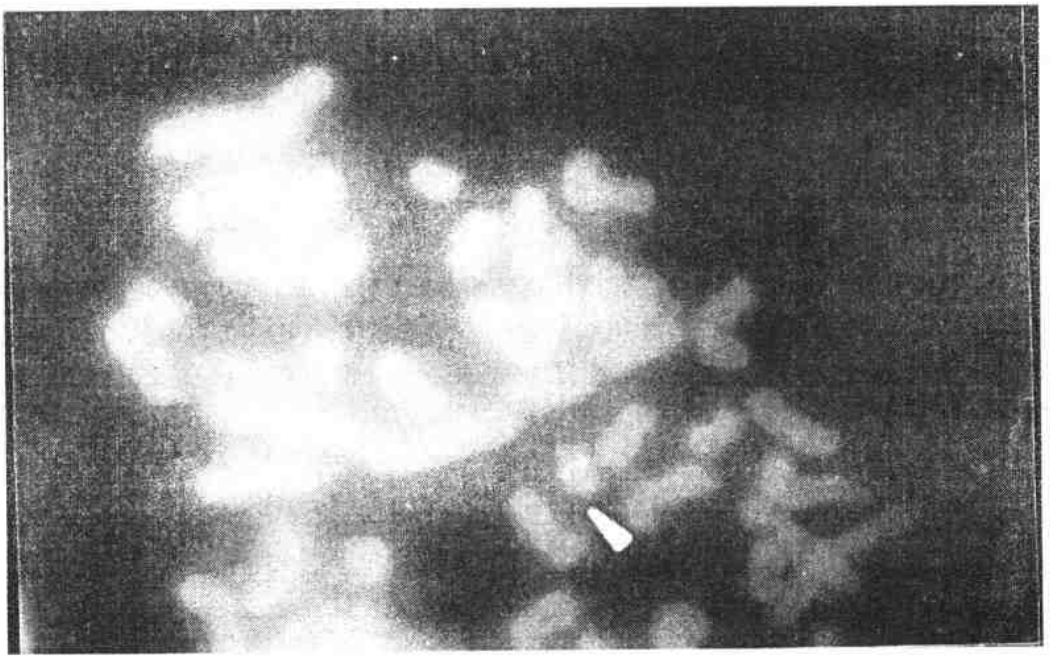


图 4－3　22q11 微缺失，仅 1 条 22 号染色体可见杂交信号(箭头所示)

第三节　FISH 技术的发展与应用

一、FISH 技术的发展

提高原位杂交的分辨率、灵敏度、容量及特异性是 FISH 技术发展的重要趋势，从单色 FISH 到多色 FISH、从中期染色体到粗线期染色体再到伸展染色质 FISH 的发展过程，使其灵敏度和分辨率由 MB 向 KB、百分距离向碱基对、多拷贝向单拷贝的方向深入。对近年来 FISH 技术的进展介绍如下：

(一)多色荧光原位杂交

FISH 的一个最大特点是可利用不同颜色的荧光分子标记不同的探针，同时对不同的靶 DNA 在同一标本上进行定位。但由于不同荧光分子的光谱重叠，一般只能同时使用 3 种不同的颜色。为了扩大 FISH 的容量，近年来发展了利用组合标记(combinant FISH)，比例标记(ratio - labelling FISH)的方法标记更多的探针。即一个探针同时利用几种不同颜色的荧光分子进行标记组合出新的颜色，理论上可标记的探针数为 2×荧光分子个数 - 1。如果调整各种荧光分子的比例，FISH 探针标记的容量将会进一步增加。

在多色 FISH 的基础上，还发展了染色体描绘(chromosome painting)，比较基因组原位杂交(comparative genomic hybridization，CGH)技术。染色体描绘是用全染色体或区域特异性探针，经多彩色荧光标记后，对染色体杂交使中期特异染色体和间期核呈现不同荧光颜色的条带，可用于识别染色体重组、断裂点分布以及鉴别染色体外核物质的起源。CGH 的基本原理是原位共杂交。将待检 DNA(如肿瘤细胞的 DNA)与相应的正常细胞 DNA 进行不同颜色的标记，混合后在两者存在竞争抑制的情况下同时对正常细胞的染色体标本杂交，可获得基因组中 DNA 扩增和缺失等的整体核型图，为遗传畸变分析提供了一种基因组指纹印记的检测手段，但 CGH 不能检测 DNA 的结构重排，而且灵敏度和分辨率受到一定限制。

(二)DNA 纤维 - FISH

在用 FISH 技术构建 DNA 分子图谱时，它的分辨率，即两个不同 DNA 探针能够被检测到的最小距离将决定分子图谱的精确程度，是最关键的技术指标。由于所使用的靶 DNA 切片的染色体或 DNA 分子结构不同，FISH 技术所能达到的分辨率水平差别很大，即染色体或 DNA 纤维的浓缩程度越高，分辨率越低。如粗线期染色体的浓缩度比中期染色体低 10 ~ 15 倍，而 DNA 纤维比中期染色体低 900 倍。下表对靶 DNA 的结构与 FISH 的分辨率，可检测 DNA 的长度及优缺点进行了比较。

表 4 - 2　在不同靶 DNA 上荧光原位杂交技术的比较

靶 DNA 结果	分辨率水　平	可检测 DNA 序列范围	优点和缺点
中期染色体	1 ~ 3 Mb	> 1 Mb	保持染色体结构，切片容易制备 分别率水平低，识别染色体困难
前期染色体	200 kb	> 200 kb	保持染色体结构 分别率水平居中，切片不易制备，识别染色体困难

续　表

靶 DNA 结果	分辨率水　平	可检测 DNA 序列范围	优点和缺点
拉长的中期染色体	200 kb	> 200 kb	保持染色体着丝点端粒结构 分辨率水平居中,拉长染色体不均匀, 识别染色体困难
粗线期染色体	100 kb	> 100 kb	保持染色体结构,识别染色体可能, 分辨率水平较高, 切片制备不易
间期细胞核	50 kb	50 kb ~ 2 Mb	分辨率水平高,切片容易制备 无染色体结构
游离染色质丝	10 kb	10 kb ~ 1 Mb	分辨率水平高 无染色体结构
DNA 纤维	1 kb	1 kb ~ 1 Mb	分辨率水平高 无染色体结构

纤维 - FISH 在人类基因组研究中的应用十分广泛，除用于基因组精确作图和辅助基因克隆外,还可定量分析不同克隆之间的排列顺序及重叠程度,决定基因之间的物理距离及方向;定量检测染色体的重排和缺失,分析靶序列的拷贝数等。纤维 - FISH 是构建邻接克隆群(Contig)图谱的有力工具。用 BAC、YAC 载体构建的 Contig 图谱需要对同一染色体上相关 Contig 之间的间隔进行封闭。现有的分子技术很难测量这些间隔的尺寸,而用高分辨率的纤维 - FISH 可以快速准确的确定克隆之间的重叠,并对重叠程度进行定量分析。

纤维 - FISH 的缺点:①DNA 纤维的随机断裂可能导致同一探针的信号长度不一致;②DNA 纤维在标本制作过程中变性程度不同,会造成局部损伤,或伸展程度不一,影响定量分析;③纤维 - FISH 的结果不能判断探针的染色体位置,应与中期染色体 FISH 结合应用。

(三)原位 PCR - FISH

即在原位 PCR 的底物中加入已被荧光标记的核苷酸,扩增后直接进行检测,省去杂交程序。其灵敏度可比常规 FISH 高 2 个数量级。它的缺点是 PCR 固有的碱基错配及由于模板损伤可在扩增杂交后产生假阳性结果。但经过条件优化和技术改进,该技术仍然很有发展前途。

二、FISH 技术的主要应用

(一)基因图谱的绘制

是 FISH 应用的主要方面,基因制图和基因定位是人类基因组计划的主要任务之一。应用 FISH 技术,对小于 1 kb 到几个 MB 的 DNA 或 cDNA 片段都可以在染色体上精确定位,而且可以使用多色探针确定两个序列间的物理距离和排序。这是现阶段其他制图方法无法比拟的。高分辨率物理图谱的获得和靶 DNA 载体的选择有关。对未知探针制图的程序一般是先进行常规的中期染色体 FISH,将探针大致定位于染色体区带水平,然后用伸展染色质进行更精确的定位,如先制备粗线期染色体找到目的基因相关的探针在特定区域的定位,确定相距 100 kb

的 DNA 序列的位置关系，进一步在 DNA 纤维上进行 FISH 操作，构建出这一区域的物理图谱。基因定位可以为基因连锁分析提供更多的 DNA 标记，反之也为更多基因的克隆提供信息。将目标基因先定位，再克隆的定位克隆策略是寻找疾病相关基因的一项重要手段，FISH 技术为将目标基因定位在染色体上提供了有效的方法。

（二）染色体数目和结构分析

高分辨 FISH 技术结合 CCD 照相机和共聚焦显微镜设备进行图像分析，可以直接观察不同成分和结构组成以及染色体数目的改变。如分析研究 Alu 序列和 LINE 序列在人中期染色体上的分布，发现 Alu 序列主要在 R 带，Linc 序列主要是 G/Q 深带。如利用 FISH 技术观察人类基因组中编码基因的分布，发现基因主要集中在 C + C 含量最丰富的 DNA 区。使用几种定位在染色单体不同部位的单拷贝探针，可判断中期染色质的螺旋构象或特定 DNA 序列在染色体骨架上的位置。此外，FISH 可以直接观察染色体端粒、定位着丝粒的重复序列以及研究减数分裂的重组过程等。

染色体 X、Y、13、18 和 21 的数量变化常与先天性疾病有关。如用染色体着丝粒重复序列作探针，很容易确定分裂细胞或间期细胞中这些染色体的数目。又如实体瘤的分裂期细胞很少，染色体研究一向比较困难。而 FISH 技术由于可进行间期细胞的数量变异分析，而且能辨认常规显带分析中难以识别的标记染色体的来源，因此在肿瘤细胞遗传学中发挥了重要作用。如 Hopman 采用 FISH 技术研究膀胱癌，发现 9 号染色体的丢失是膀胱癌早期恶变的标志，而且随着疾病的进展，伴随有额外 1 号染色体的出现及四倍体的改变。

FISH 技术还可检测间期和中期细胞的染色体易位及位置特异的断裂点。在正常的二倍体细胞，每一对染色体具有一个独立的荧光信号，而易位的染色体上显示两个不同的荧光。染色体彩绘技术尤其适于这种易位分析。

（三）基因扩增与缺失的检测

在 FISH 技术之前采用的检测基因扩增的分子生物学方法如 Southern blot 等不仅费时费力，而且不可能在细胞水平上观察到基因扩增的状态。FISH 特殊的分辨力和灵敏性使其能定位亲本基因和扩增基因，在间期细胞核上显示 DNA 扩增的直接证据，而且根据扩增 DNA 荧光信号的数量和荧光强度可以判断 DNA 扩增的水平。特别是当细胞遗传学发现有异常的显带区域或异染色质区时，可根据其来源和位置用 FISH 技术进一步得到直接清晰的扩增信息。同样，根据荧光信号的有无也可以检测缺失。在肿瘤和许多遗传性疾病的研究中广为应用。

FISH 作为一种方法学对于染色体研究具有巨大的发展潜力，如活细胞 FISH，FISH 结合蛋白检测以及利用高分辨纤维 - FISH 研究细胞核骨架与基因表达的关系等，它可以为遗传重组和基因调控的研究奠定结构基础。

（曹慧青　杨菊先）

第五章

辐射性杂交

内容提要

本章主要以 G3 嵌板为例介绍了辐射性杂交的基本原理和在 EST 定位、寻找新基因、基因克隆、染色体上位点排列及测距等方面的应用，以及与 RH 技术有关的网站和数据库。

辐射性杂交(RH)技术是继荧光原位杂交技术后，国际上最常用的基因定位的方法之一。RH 作图法是指依靠辐射引起断点，利用类似于遗传重组原理和最大似然性的统计学方法，计算存在于 DNA 片段上的多态性或非多态性标记之间的断点频率，估计标记之间的距离。

第一节　辐射性杂交的基本原理

辐射性杂交(radiation hybrid，RH)制图技术是 1975 年由 Goss 和 Harris 创立的一种体细胞杂交技术，适用于构建人类基因组长范围内的高分辨率连续物理图谱。成熟的辐射性杂交制图技术是由 Cox VR 等人于 1990 年建立的。利用高剂量的 X 射线将候选染色体打断成若干片段，含有这种片段的细胞可与仓鼠细胞形成杂交克隆。在这种杂交中，人类染色体片段被插入到仓鼠染色体的中间部分，因此大部分克隆片段在进行有丝分裂时处于稳定的状态。利用类似于遗传重组原理和最大似然性的统计学方法来计算存在于 DNA 片段上的多态性或非多态性标记之间的断点频率，以此估计标记之间的距离，并建立人类基因组拷贝库—体细胞杂交系统。可根据不同要求构建出断裂程度不同的多套拷贝，从而得到不同分辨率的放射杂交图。总之，辐射性杂交是建立在受到 X 射线照射的供体细胞与非放射线处理的受体细胞融合基础上的一种体细胞遗传学方法，是利用两个标记之间被 X 射线打断的概率来计算标记之间距离，通过类似于减数分裂连锁制图的方法来确定标记之间的顺序。

通常在 RH 作图中所用的标记是序列标签位置(sequence - tagged site，STS)。STSs 作为一种基因组染色体路标是由 Olson 等于 1989 年提出的，是物理构图的一次革命。STSs 是在染色体位置已定的、核苷酸序列已知的、且在基因组中只有一份拷贝的 DNA 短片段，一般长 200 ~ 500 bp。STSs 可来自随机的 DNA 序列、以前已经定位的遗传学标记和表达序列标签(expressed sequence tag，EST)。STSs 是从单拷贝 DNA 序列发展而来的一对 PCR 引物，因此 STSs 是基于 PCR 方法的 DNA 标志，操作简单快速。每个 STSs 在基因组内有特定而单一的同源位置。特定的 STSs 在一套基因组结构上只出现一次，使得把相应的克隆库中的克隆进行排序变得更准

确、更科学，STSs 在物理图谱中的重要性有：

1. 它们可用来特异性的定义 YAC、粘粒或噬菌体克隆。

2. STSs 可鉴定出与特定克隆存在重叠的克隆。

3. 在计算机数据库中的各种物理图谱可以用 STSs 这种通用语言统一起来。

基于 STSs 的物理图谱不仅对各种方法构建的物理图谱进行整合，而且也可对遗传图谱、基因图谱等各类图谱进行整合。因此，在基因组作图和测序中，各个实验室发表的 DNA 测序数据或构建的物理图可用 STSs 加以鉴定和验证，并确定这些测序的 DNA 片段在染色体上的位置。用 STSs 作为通用的标记，有利于汇集各实验室发表的数据和资料，保证作图和测序的准确性，并把用不同方法获得的图谱统一起来实现资源共享。

一、构建 RH 图谱的方法

1. G3 嵌板的产生。用 10 000 Rads X 射线照射人类男性二倍体淋巴母细胞系 RM，然后在 50%(m/v)PEG 存在的条件下与中国仓鼠细胞系 A3 融合。融合细胞用 HAT 选择性培养 10～14 d。分离培养独立的克隆细胞，抽提基因组 DNA。

2. 确定 STSs。用 Primer 软件设计 PCR 引物，选择适合多数 STSs 扩增和 PCR 产物大小为 100～300 bp 的 PCR 条件。

3. PCR 体系及反应条件。对于每一个 STS 来说是由 83 个辐射性杂交系、2 个人类基因组 DNA 作阳性对照、2 个仓鼠基因组 DNA 作阴性对照、9 孔以水为对照组成的 96 孔 PCR 模式。

反应条件：

94 ℃ 90 s → 94 ℃ 15 s → 62 ℃ 23 s → 72 ℃ 30 s → 72 ℃ 210 s

↑ 30 次循环

PCR 扩增后，每个反应体系中加入 5 μL 上样缓冲液，总体积为 15 μL。

取 12 μL 的混合液点样于 3% 琼脂糖胶，1 × TBE，EB 染色。*Msp*I 酶切的 pBR322 作为分子大小的标记。

电泳后紫外线照射，640 × 480 – pixel CCD 照相机拍照。将图像传送到 UNIX 工作平台。

为避免出错，每一个 STSs 都要平行做两次 96 孔模式的 PCR。因此，每一个 STSs 需做 192 次 PCR，走两块电泳胶。

用 DNA/GUI 软件对 PCR 结果进行半自动分析。若一个 STSs 于同一个杂交系中，并在两块胶中都有阳性条带或都无条带，则对该 STSs 在相应杂交系进行得分计算。但是，若在两块胶中的结果不一致，这样的 STSs 不易使用。平均每一个 STSs 在两块胶中的不一致率应小于 1%。若一个 STSs 不一致率大于 8%，则放弃。用该方法 G3 RH 嵌板共对 8 196 个 STSs 得分。

4. 构建 RH 图谱。构建 RH 图谱是利用从人类双倍体细胞全基因组 RHs 获得的资料，使用计算机软件—SAMAPPER，构建含许多标记的 RH 图。SAMAPPER 是利用标准的最大似然性合计临近标记间的匹配可能性，从而估计某一个标记的位置。G3 构图法是首先初步确定在某一染色体上所有标记的最大可能性顺序，然后分成 20～35 个连锁群，同时确定在每一个连锁群中标记的可能顺序。为实现长距离的连续性和确定连锁群间的相对顺序，G3 构图法还利用了 Genethon 减数分裂连锁图中已排好序的 1 381 个 STSs，以模拟退火法(simulated annealing algorithm)来研究标记最可能的位置及标记间的顺序。所以这又称为综合排序(comprehensive or-

der)或高可信度的排序。

在上述 8 196 个 STSs 中,有 1 181 个在 RHs 中的得分模式与一个或更多个 STSs 相同,像这样的 STSs 称为重复 STSs。因此,只有 7 015 个 STSs 可被用作 RH 作图。在同一连锁群中的所有标记,至少与该连锁群中的一个标记的 LOD 值等于或大于 6。对于某一条染色体上用至少含一个 Genethon STSs 的 LOD 6 连锁群作图,7 015 个 STSs 中同时符合上述条件的只有 5 545 个。然后每一个染色体的所有连锁群形成一个连锁群。最后利用模拟退火 algorithm 对每一个染色体上的标记进行排序。总之,该过程产生了独特性的连锁群,利用 Genethon 减数分裂连锁图中已排好序的标记来识别和纠正连锁群间错误排序,实现了图谱长距离的连续性。

一条染色体综合排序和 1 000:1 bins 系建立后,每次从所建染色体图谱中删除一个标记,计算该条染色体的总长度。如果删除一个标记,使染色体图谱长度缩减 > 20 cR 10 000,则该标记称为扩增子(expander)并将永远从图中删除,那么该染色体上的高可信度标记顺序和 1 000:1 bins 系又具有新的连锁关系,需重新计算并寻找扩增子。每条染色体至少经过四次循环来删除所有的扩增子。在 5 545 个用来构建图谱的 STSs 中,有 496 个扩增子。因此在整个 RH 图谱中共有 5 049 个排序的 STSs 和 1 766 个 1 000:1 bins。除此以外,945 个重复 STSs 与上述 5 049 个 STSs 的某一个具有相同的得分模式,所以也可以说 RH 图谱中对 5 944 个 STSs 进行了排序和定位。

二、RH 嵌板

目前关于人类基因组 RH 图谱的嵌板有 3 种:用 3 000 rad X 射线处理的 Gene Bridge 4 (GB4) 嵌板、用 10 000 rad X 射线处理的 G3 嵌板、用 50 000rad X 射线处理的 TNG 嵌板。其中最常用的有 GB4 和 G3 嵌板。GB4 嵌板是由人成纤维细胞 HFL 121 作为供体,中国仓鼠细胞 A23 作为受体进行融合,并经较低能量 X 射线(3 000 rad)照射产生的,由 93 个细胞株组成,这个系统可用于未知 DNA 标志的定位。DNA 标记之间的定位精确度约为 1 ~ 5 Mb。该系统可以与连锁分析一样测定大区间(0.3% ~ 1%基因组)内的连锁关系。这种大跨度连锁关系(称为"自上而下"的信息)的确对于染色体或基因组范围的连锁图谱或物理图谱的构建起着重要作用。G3 嵌板是以人类 RM 淋巴母细胞株为供体,仓鼠胸腺嘧啶激酶缺陷 A3 细胞株作为受体,经细胞融合并经 X 射线 8 000 rad 照射产生的,共包含 83 个不同的杂种细胞株。应用这一系统可以测定不同 DNA 标志在染色体上的次序并给出标志之间的相对距离,其精确度可达 100 kb。该系统适用于对某一 DNA 已知其粗略染色体位置而需要更精细定位时使用,而不适用于未知 DNA 标志。对未知标志需要先用其他方法(如用 GB4 嵌板或 FISH 法)进行粗略的定位。

表 5 - 1　3 种 RH 图比较

嵌板	GB4	G3	TNG
嵌板中杂交系数目	93 + 2 个对照 (HFL 和 A23)	83 + 2 个对照 (RM 和 A3)	90 + 2 个对照 (MR 和 A23)
每个杂交系存留染色体(%)	30%	18%	16%
片段的平均大小	10 Mb	4 Mb	800 kb
X 射线剂量与距离	1 cR_{3000} = 270 kb	1 cR_{10000} = 25 kb	1 cR_{50000} = 4 kb
RH 图谱的平均分辨率	1 000 kb	267 kb	60 kb

第二节　RH 图谱的应用

辐射性杂交技术是继荧光原位杂交后新近建立的染色体定位方法，RH 作图法提供了一种联系物理图和遗传图的方法，已成为当今构建人类基因组大尺度、高密度、连续的染色体图的常用方法之一。其用途主要有：EST 定位、基因克隆、基因组作图、测定距离、寻找新基因等。

一、EST 定 位

这是 RH 目前最主要的作用。RH 对 EST 定位快速、简单易行、定位精度高，根据统计学方法可直接定位。到目前为止，新发现的基因用 RH 定位的有上百条。现以 G3 嵌板为例，以 RH 图对 EST 进行定位的过程。

1. 引物设计。根据所要定位基因的序列设计引物。引物长度 25 ~ 30 bp 之间，退火温度 Tm > 65 ℃，两条引物之间的 Tm 差值≤2 ℃。

引物的位置要求位于 3′非翻译区，因为人类基因的 3′非翻译区很少有内含子，这样从基因的 cDNA 估计的 PCR 产物的长度与从基因组 DNA 扩增的长度相同。引物设计好之后，应作核酸同源性比较，应避免引物与仓鼠的基因组的同源性。如无法避免，则在引物 3′端至少有 3 个碱基与仓鼠的基因组有差异，这是因为 RH 定位的模板是 83 个人与仓鼠的杂交细胞系。

2. PCR 反应体系：

MilliQ H_2O	5.6 μL
10 × 反应缓冲液	1.0 μL
5′引物(10 μM)	1.0 μL
3′引物(10 μM)	1.0 μL
dNTP(200 μM each)	0.2 μL
G3 放射杂交系	1.0 μL
Elongase (*Taq* 酶)	0.2 μL
总体积	10.0 μL

3. 循环参数：

温度	时间(min)	
95 ℃	10:00	
94 ℃	0:30	30 cycles
60 ℃	0:30	
72 ℃	0:23	
72 ℃	3:30	
4 ℃	forever	

4. 电泳。PCR 反应后，每个反应体系中加入 2 μL 3 × loading buffer

3 × loading 染色液：

0.1M EDTA pH 8.0

0.025%(w/v) 溴酚兰

10%(w/v) Ficoll 400

制备 3% 的 Agarose 胶，插上 4 个 26 齿梳子，放入含 1 × TBE 电泳液中。每个泳道上样 10.0 μL反应液。在每一排的最外面的泳道中加入用 *Msp* 1 消化的 pBR322 marker。在电压为

120 V 条件下，电泳 45 min。

5. 电泳图像。将电泳胶放入紫外线暗箱中，用 640 × 480 的 pixel CCD 照相机照相，并以 TIFF 格式保存下来。

6. 分析。在结果中有 PCR 产物的杂交系记为 1，没有的记为 0，因此，该结果是由一系列的 1 和 0 组成。利用软件自行分析或将结果及电泳图 E－mail 到 RH 服务中心，只需等待 1～2 个星期就可获得定位结果了。G3 嵌板和 Cene Bridge panel 的分析软件可以从互联网（http://www.sph.umich.edu/group/stagen/software）获得或从 Mike Boehnke 博士（Beohnke@sph.unmich.edu）处获得；也可把 PCR 结果数据通过 E—mail 分别送往 Stanford（rhserver@shgc.stanford.edu，使用 G3 嵌板时）或 Whitehead/MIT Center for Genome Research（http://www－genome.wi.mi.edu/cgi－bin/contig/rhmapper.pl，使用 GB4 嵌板时），数据就会被自动分析并把结果在数小时内送回。现在绝大多数研究者是采用第二种方法。

在定位时可联合使用几种嵌板，如 GB4 和 G3 嵌板联合使用，对 G3 嵌板无法定位的，也可用 G4 嵌板。也可与其他技术（如荧光原位杂交技术）联合使用。现以纤维母细胞生长因子 18（FGF18）为例阐述如下。

以 FGF18 cDNA 的序列设计引物：5′GAC GAT GTG AGC CGT AAG CAG CT 3′和 5′GCA TAC TTG TCC CCA TCC TCG CCG CG 3′。以 GB4 和 G3 嵌板为模板经过上述①～⑥步骤结果分别为 10000 10100 11001 00000 00001 10010 00010 00100 01000 11000 10000 0000100010 00000 00000 00000 01010 00011 001 和 10000 10000 00001 01000 00000 00000 00010 00000 00000 00000 10000 00000 01020 10000 00000 00000 000。将结果分别 E－mail 到 Whitehead/MIT Center for Genome Research（http：//www－genome.wi.mi.edu/cgi－bin /contig/ rhmapper.pl）和 Stanford（ rhserver@shgc.stanford.edu）。以 GB4 为模板定位结果是距 WI－6737 标记的远端 0.10cR3000（LOD = 15），以 G3 为模板定位结果是距 SHGC－14309 标记的远端 8cR10000（LOD =8.76）。这两个嵌板的定位结果都是将 FGF18 定位在 NCBI GeneMap '99 上 D5S504（175.3 cM） D5S677（182.4 cM）之间，即定位在 5q34～q35 染色体区段间。

人 StanfordG3 和 GeneBridge4 嵌板的建立，加强了不同 RH 嵌板之间的联合，加速了新 EST 的定位。EST 被组装进有重叠的转录单位，和遗传图的多态标记一起排序，产生的不同基因组作图来源的综合标记图，对其基因组的遗传图、物理图和转录图完成有重要作用。

二、寻找新基因

利用 RH 图中的标记，可寻找其周围的基因。例如，染色体的 2p15－p6 区位于多态标记 D2S119 和 D2S378 之间，包含约 16cM 的距离，是未知盲区，利用构建的 2p5～p16 RH 图发现有 2 个基因 CNC 和 DHRD 位于该区域内。

Horrigan 等构建了一个 6Mb 间隔的人 5 号染色体 q31 区的高分辨率图，包括 175 个 STSs；其中 33 个有遗传多态性。该图促进了几个基因的发现，包括躯体肌肉营养失调症（limb girdle muscular dystrophy type A）和急性髓性白血病和骨髓发育不良症（acute myeloid leukemias and myelodysplasia）的基因。显示利用 RH 嵌板能有效地确认新基因。

三、基 因 克 隆

20 世纪 70 年代开始的人类基因克隆主要采用“功能克隆法”，即从相应的 mRNA 或蛋白质出发，借助免疫反应分离纯化相应的 mRNA 或蛋白进行氨基酸测序，然后以此为探针克隆筛

选出所要的基因。20 世纪 80 年代开始借助物理图谱尤其是 RH 图谱进行“定位克隆”。其基本思路见图 5－1。

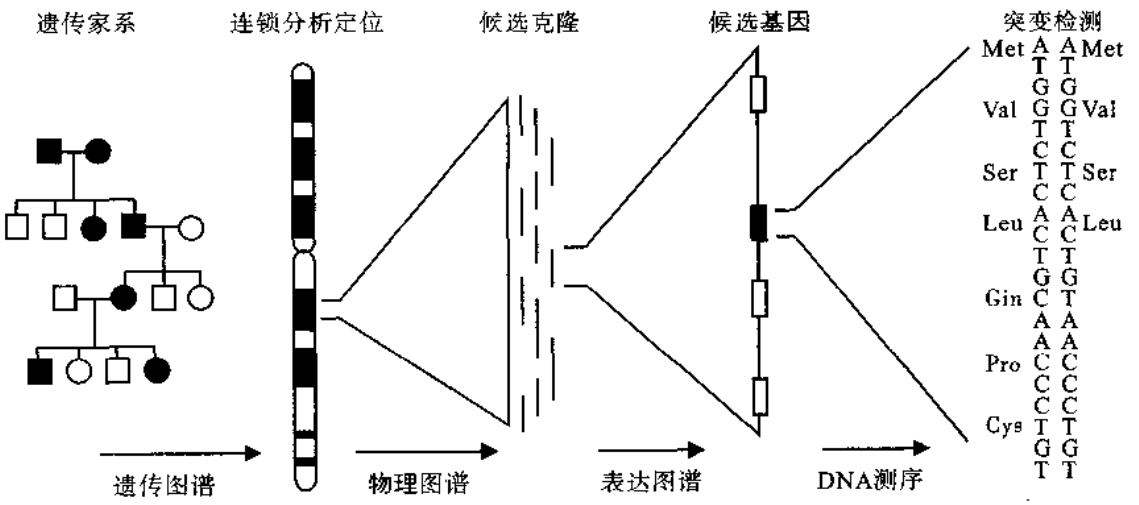


图 5－1　人类物理图谱在疾病基因克隆研究中的应用

随着基因组测序的完成，表达谱的发展，EST、cDNA 或全长 cDNA 资料的日益丰富，一旦某一基因被定位，很容易从数据库中查到可以作为候选的基因或 DNA 顺序并可立刻进行突变检测，这就给基因的识别和克隆创造了条件。

四、测 定 距 离

RH 图是检查染色体上位点排列和估算它们之间距离的一种有效方法。RH 图含有大量位置和排序高度忠实性的遗传标记，经过相应的统计软件分析标记之间的连锁关系，可得到标记和 EST 之间的距离，对于位置和排序不确定的少量标记，标记的位置可根据多数情况下的表现进行估计。

五、RH 相关网站及 RH 数据库

要想更加详细的了解 RH，可查询 RH 相关网站和 RH 数据库。

RH 相关网站：

http://compgen.rutgers.edu/rhmap

http://www.ncbi.nlm.nih.gov/entrez/genome/org.html

http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper.pl

http://www-shgc.stanford.edu

http://www.ebi.ac.uk/rhdb

RH 数据库是储存与辐射性杂交相关的原始资料的数据库，包括关于嵌板、实验条件、STSs、实验结果、RH 图谱、作者及文献目录等。研究者可通过以下几种方式查询 RH 数据库：

Internet:　http://www.ebi.ac.uk/RHdb

FTP server:　ftp://ftp.ebi.ac.uk/pub/databases/RHdb

Email:　　rhdb@ebi.ac.uk(enquiries and submission)

目前 RH 数据库中已含有几千个 DNA 位标,对于任何未知 DNA 标志的定位与其他方法相比是非常简便的,可在一天内得到结果。

六、RH 的优缺点

荧光原位杂交(FISH)法和辐射杂种细胞系(RH)技术是国际上最常用的基因定位的两种方法,各有优缺点。FISH 可以定位基因组中多个同源位点,结果直观、可靠,而 RH 法则很困难。但是 FISH 法检测步骤繁杂,尤其受探针大小的影响较大,2kb 以下的 cDNA 序列很难定位,而且结果需要有经验的细胞遗传学家进行分析。RH 法简单易行,定位精度高,适合小于 1kb～2kb 以下的序列,根据统计学方法可直接定位。其结果也便于不同实验室间比较。和其他作图法相比,RH 作图法是依赖于辐射引起的断点而不是靠减数分裂重组得到的断点,是一种完全独立的方法,所以是一个互补的作图方法,可利用所有不同的 STSs,具有巨大的 DNA 补充潜力,从而有利于整合不同的遗传和转录图以及所有基因组位标(无名序列、中心粒和端粒)。

但是 RH 法也受到某些自身条件的限制,RH 嵌板中每个杂交克隆均包括仓鼠和人的基因组 DNA,可以看作是在仓鼠基因组 DNA"背景"下相区别。如有些基因无法定位的原因即在于它的序列包括阅读框架以及 3′UTR 区域与仓鼠的相似性高达 90%以上。此外,做 ESTs 或全长 cDNA 定位的引物为 cDNA 序列,而 RH 法是在基因组 DNA 上扩增出相应的片段,由于内含子序列的存在,扩增结果与引物设计的部位直接有关,一般应选用 3′UTR 部位的序列。还应指出,用 RH 法能否成功定位,依赖于该基因所在染色体区域上合适位标的存在。如果现有的 RH 图中此区域分布的合适位标过少,可改用位标位点多的嵌板。因此随着今后 RH 图精度的提高,越来越多的 STSs 被放置于染色体上,有望弥补这个缺陷。

(孟宪芳)

第六章

计算机克隆及染色体定位

内容提要

简要介绍了计算机克隆的基本原理、背景和具体操作步骤；用计算机分析进行染色体的精确定位；并简单提示熟练掌握和运用美国生物信息中心(NCBI)的各种数据库和软件可提高基因分析的效率。

第一节　计算机克隆

一、基本原理

传统上，克隆一新基因的 cDNA 全长涉及文库的筛选，因而很费时费力。随着人类基因组计划测序的完成和转录图计划的进行，现在借助计算机可大大加快基因克隆的过程。计算机克隆主要是借助计算机和 ESTs(Expression sequence tags)分析从而克隆基因。ESTs 是短的(约 200～600 bp)、源于 cDNA 克隆 3′或 5′端的、只经一次测序的 cDNA 序列片段。由于 ESTs 数据库中有大量来源于各种生物(截至到 2002 年 6 月 28 日，GenBank 中共有 11 948 269 条 ESTs，其中有 4 464 710 条人 ESTs。)、各种不同发育阶段和不同组织来源的 ESTs，所以同一个基因会有许多相互重叠的 ESTs 序列，因此可以采用头尾相接的方法不断延伸从而用其来构建一 contig(重叠克隆或序列的拼接即为 contig，这里译为“重叠克隆群”)，再以该重叠克隆群为被检序列，重复相似性比对延伸过程，直至不能延伸为止。这一重叠克隆群有可能代表了一基因的 cDNA(全长)。在这么多的 ESTs 中，通过仔细的筛选，有可能找出一个基因的大部分序列，这就是计算机克隆的大致原理。

二、现状与发展

早在 6～7 年前已有文献报道了可用计算机克隆出新基因，国内近 1～2 年亦有这方面的报道。然而由于 ESTs 序列信息的快速增长，使用 ESTs 数据库克隆新基因或蛋白家族新成员变得越发困难。如从已知的 100 个左右的小 GTPase 家族出发，寻找该家族新成员，通常过程如下：这 100 个已知序列必须首先搜寻约 4.5 百万条 ESTs(截至到 2002 年 6 月 28 日)，每次数据库搜寻平均产生 500 条与小 GTPase 有显著相似性的 ESTs，最后的工作是将这 100 次搜寻所

得 ESTs 与已知蛋白序列比较，以寻找与原 100 个小 GTPase 家族成员不同的序列。若考虑到 ESTs 测序的错误和同源序列的进化差异，这一过程则更为复杂。目前有一些计算机程序可用以辅助分析这一过程（如 ESTblast，其网址为 http://www.hgmp.mrc.ac.uk/ESTBlast/；E-CLONE，其网址为 http://bioinfo.biosino.org:9090/clone.html 等试图通过计算机直接克隆基因的 cDNA 全长。），并有一些新方案。但是用 ESTs 拼接的方法并非总能得到理想结果。作者认为，ESTs 对绘制人类遗传图和物理图起了很重要的作用。但是随着人类基因组序列图的完成和转录图的进行，单纯用 ESTs 拼接的方法克隆新基因的价值已远不如几年前。国外高水平的科学杂志上很难见到对人类单纯用计算机克隆新基因并进行功能研究的论文（其他物种除外）。但是 ESTs 在人类基因组研究中继续扮演着其他重要角色，例如可以用 ESTs 估计基因的时空特异性分布，从而解析基因的功能。当然由于数据递交的偏性，用计算机估计的结果必须用实验方法如经典的 Northern Blot 验证。

三、操作步骤

根据 21 和 22 号染色体上的基因结合 ESTs 估计，人类共有约 30 000～35 000 个基因。据人类基因组测序联合组创建的 IGI.1（initial integrated gene index）估计，目前含 24 500 个真正的基因，还有数千个人类未知基因。在这些未知基因中，属全新的基因（即不属任何已知家族的基因）非常少，许多基因可能属已知家族的未知成员，因而现在很少有人会漫无目的地从浩瀚的 ESTs 数据库中寻找新基因，因此借助计算机分析，在克隆已知家族新成员方面尚有作为。

然而由于人类基因转录调控机制的复杂性，在实际克隆新基因的过程中比较复杂，例如人的 UniGene 数据库（由 GenBank 通过计算机分析衍生而来）共有 101 698 簇（UniGene Clusters），其中含 632 551 条未知 ESTs（UniGene Build #151，2002/5/27）。而据 UniGene 的定义，每一 UniGene 簇应对应于一个基因。这里想说明的是，任何借助计算机克隆的结果必须得到实验的验证。

用计算机克隆已知家族新成员的一般步骤如下：

1. 以已知家族成员保守的氨基酸序列为探针，查询 ESTs 数据库。

2. 仔细筛查相匹配的 ESTs，以决定哪些是已知家族基因的 ESTs，哪些可能代表了未知家族成员的新 ESTs。

3. 利用 ESTs 具有相互重叠特性，尽量从数据库中搜寻出可能代表新成员的 ESTs。

4. 用计算机将搜寻到的所有 ESTs 进行拼接，产生最长的 cDNA 序列，并比较其与已知基因的关系。

5. 计算机分析所产生的最长的 cDNA 序列有无完整读码框，根据不同组织来源的 ESTs 估计该“基因”的组织分布等。

6. 实验验证：包括 RT-PCR（有时 5′RACE）验证全长 cDNA。Northern Blot 证实长度、选择性剪接转录本及组织学分布。

四、一点经验

正如上面所说，实际工作中用计算机克隆涉及多方面的知识，这里简单介绍笔者的一点体会，可供参考，希能对读者有所裨益。

1. 由于 ESTs 是一次测序结果，可能有测序错误，所以如拼接的序列不含较长的读码框时，此时要结合核酸相似性比较的结果，仔细鉴别其是否确为测序错误。

2. 许多情况下用 ESTs 拼接所产生的重叠克隆群并非 cDNA 全长，这里有建库的原因。基因的真正外显子常处于重叠克隆群的前一个外显子。此时若能结合 Northern Blot 结果，外显子中等大小的长度约为 122 bp，平均大小为 145 bp，据此可以推断其真正的 cDNA 全长。

3. BLAST 搜索的 *E* 值对最后结果很重要。*E* 值又称期望值(expectation value)，是指在等于或大于分值 *S* 时，搜索到的排列在数据库中随机发生的可能性。*E* 值越低，得分 *S* 越有显著性。BLAST 搜索证实一致序列时，*E* 值太高，会导致匹配旁系同源物(见本书计算机分析部分)；太低(即太严谨)会导致假阴性，即由于 ESTs 测序错误等原因搜索不到真正阳性 ESTs 序列。实际操作中对 ESTs 数据库可将 *E* 值定在 10 ~ 20 e 左右。实际操作时需经 BLAST 搜索预实验，定 *E* 值范围。

4. 寻找新基因时，若结合基因预测软件(如 GenScan：http://genes.mit.edu/GENSCAN.html)和其他物种的同源基因信息，有时能提供相互参考的信息。

5. 利用 UniGene 的 ESTs 和 SAGE(http://www.ncbi.nlm.nih.gov/SAGE/)进行组织学分布不是特别可靠。例如心脏的 Troponin I(cTnI)是一个已知的心脏特异表达的基因，其与 Troponin T 和 C 构成横纹肌 Troponin 复合物。cTnI 的 UniGene 号为 Hs.351582 ，该 UniGene 簇所包含的所有 ESTs 中，除了大部分为源于心脏的 ESTs 之外，还有源于睾丸、肺、子宫、卵巢和神经母细胞瘤等组织的 ESTs，但是并不能据此得出 cTnI 在这些非心脏组织中表达的结论。

6. 利用 ESTs 克隆新基因，最好直接从 ESTs 数据库中挖掘数据，而不要利用 UniGene，其中一个原因是 UniGene 数据库包含了许多不是特别可信的信息。

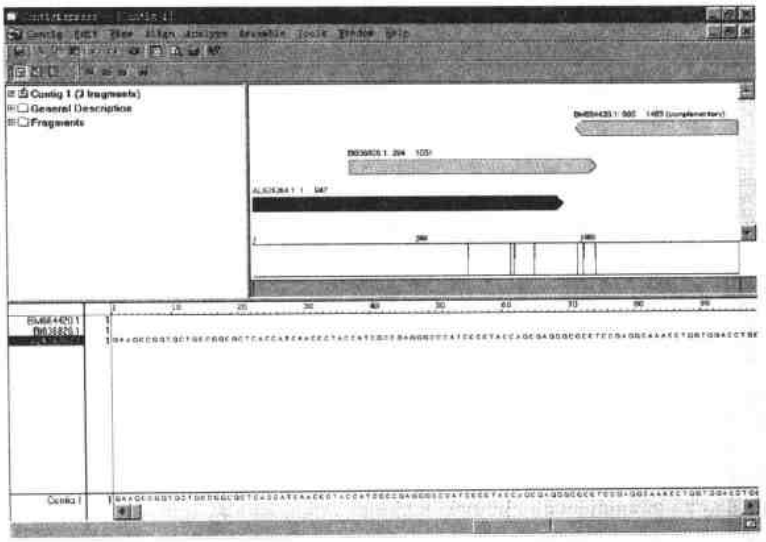
7. 在一给定文库中代表同一基因的 ESTs 数目的多少可作为来源于该文库的某个组织特定基因表达水平的粗略指标，有时侯可以利用多种文库总的匹配的 ESTs 大概估计组织学分布。但是切不要据此当作真正的组织学分布结果，因为数据库中某一组织文库可能非常小而导致在该组织表达的基因的 ESTs 不被检出来，而相反，有些组织有多个文库，如 UniGene 含 23 个心脏 cDNA 文库，而源于鼻组织的文库有 2 个。这就在一定程度上造成数据递交的偏性。

五、简单实例

下面给出一个用 Vector NTI 软件包进行 ESTs 序列拼接的简单例子。例如已知 3 条 ESTs 序列，其 GenBank 接受号分别为 BI836826.1 、BM664420.1 和 AL525264.1，为了将其组装成一全长 cDNA，其大致过程如下：

首先从 GenBank 下载这 3 条 ESTs 序列至本地 Vector NTI 数据库。随后打开数据库，选定这 3 条 ESTs(图中阴影部分)，同时用 Assemble 程序中的 ContigExpress 命令进行组装。

经过计算机简单拼接，这 3 条 ESTs 被组装为一个重叠克隆群。图中最下方的 contig 1 即是拼接完成的 cDNA。



第二节　染色体定位

在传统上，基因的染色体定位可以有家系分析（如经典的红色盲基因定位）、用体细胞遗传学和分子生物学（如 FISH）分析等。然而随着人类基因组测序计划的完成，利用计算机进行基因的染色体定位则变得简单易行。

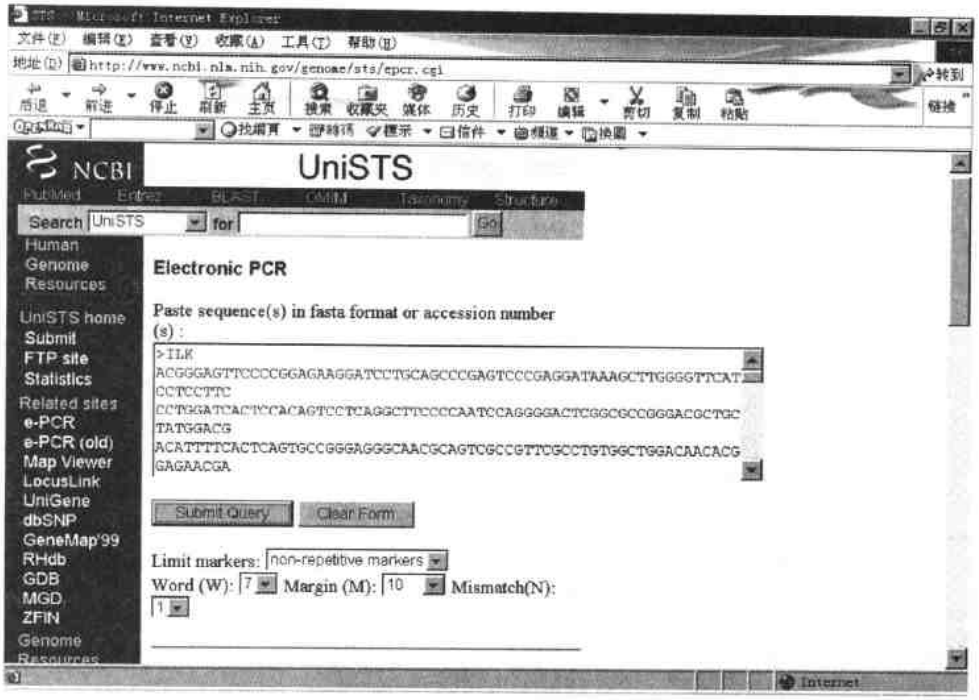
人类基因图有细胞遗传学图（基因定位于染色体区带）、基因连锁图（基因间的连锁关系和相对距离，以 cM 为单位，cM（centimorgan）：厘摩，遗传图上两基因间的距离为 1 cM 时，相当于约 1 000 kb）和物理图（基因组 DNA 的序列排列）等。本文染色体定位指细胞遗传学图和物理图而言。

目前利用计算机分析，可以有很多种方法对任何一个基因进行染色体的精确定位。这里介绍用美国国立生物技术信息中心（NCBI）数据库进行基因定位的几种方法。

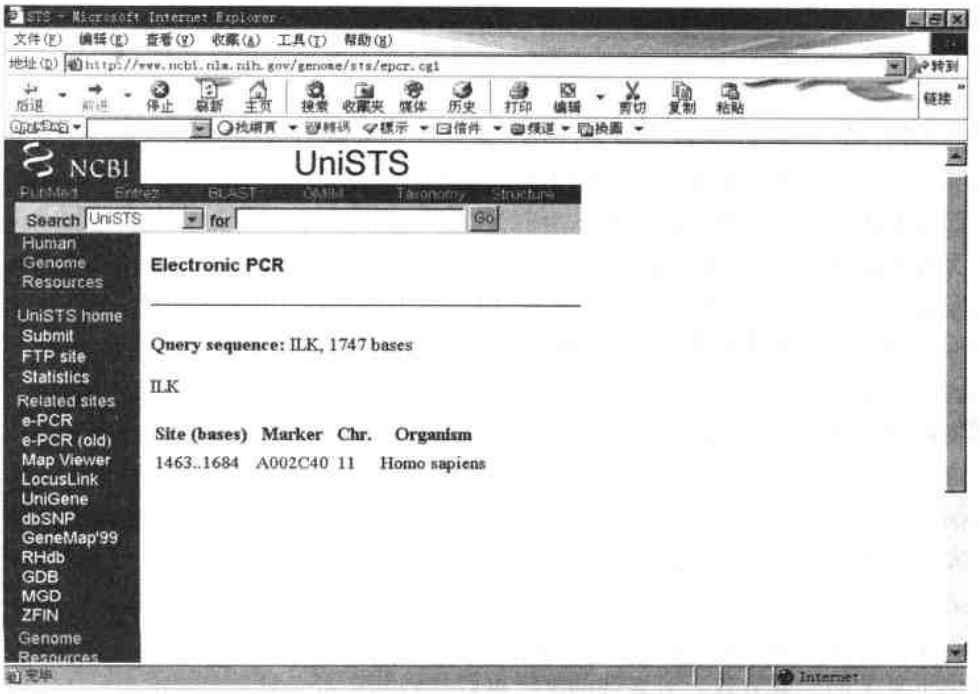
若已知基因序列，可以通过 e－PCR 寻找 STSs（Sequence tagged site）或者直接用 BLAST 将序列对人类基因组进行同源比对（http://www.ncbi.nlm.nih.gov/genome/seq/HsBlast.html）。STSs 是序列及定位已知的、长度在 200～500 碱基并且在整个人类基因组中只出现一次的基因组序列。STSs 可作为人类基因组物理图的界标，并且可用于基因的定位。STSs 序列易被 PCR 扩增和检测，在 STSs 数据库中包含了引物和 PCR 条件。ESTs 即是来源于 cDNA 的 STSs。人类基因组计划物理图分辨率原定目标是 100 kb/1 个 STSs（人类基因组长度以 3×10^9 bp 计，共需约 30 000 STSs，然而根据 NCBI 的基因组研究统计表明（http://www.ncbi.nlm.nih.gov/genome/guide/human/HsStats.html），到目前为止一共产生了 113 292 STSs，所以平均下来目前物理图的分辨率大约在 25 kb/1 个 STSs。这足以保证在很多情况下可直接通过 STSs 进行基因定位。

若不知基因序列，只知道基因名称，可以通过人类基因组图浏览器（Human map viewer，http://www.ncbi.nlm.nih.gov/cgi－bin/Entrez/map-search）进行关键词查询。该主页的高级查询则提供了复杂的检索方式。有兴趣的读者可以登陆其主页一试。

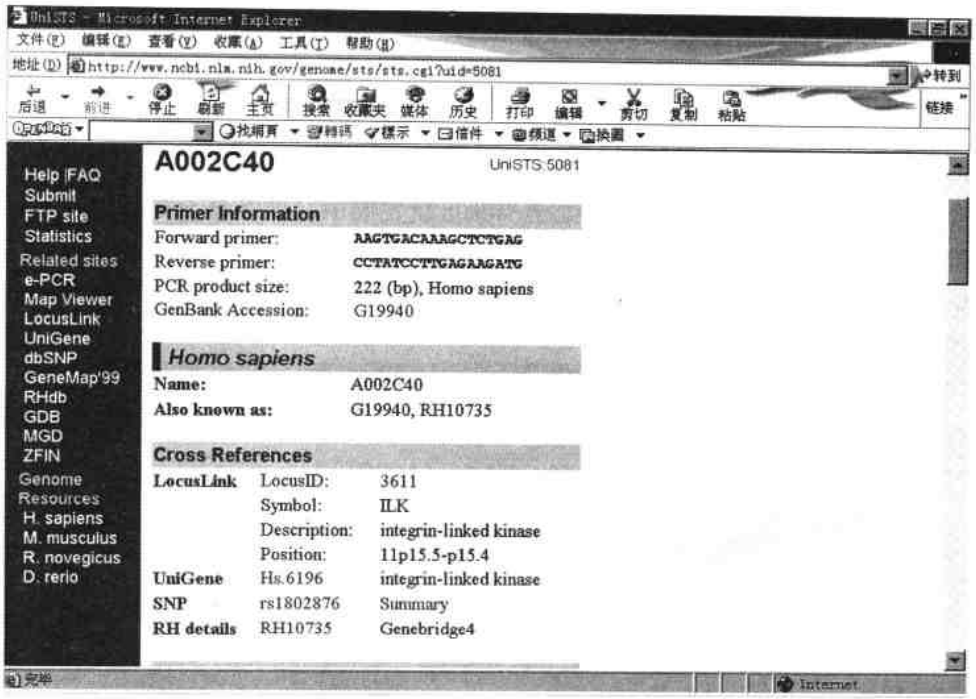
例如:已知人整合素连接激酶(ILK,GenBank 接受号 20561096)的 cDNA 序列,预查询其染色体定位。进入 Electronic PCR 主页后(http://www.ncbi.nlm.nih.gov/genome/sts/epcr.cgi),将 ILK 激酶的 cDNA 序列输入,点击发送即可。



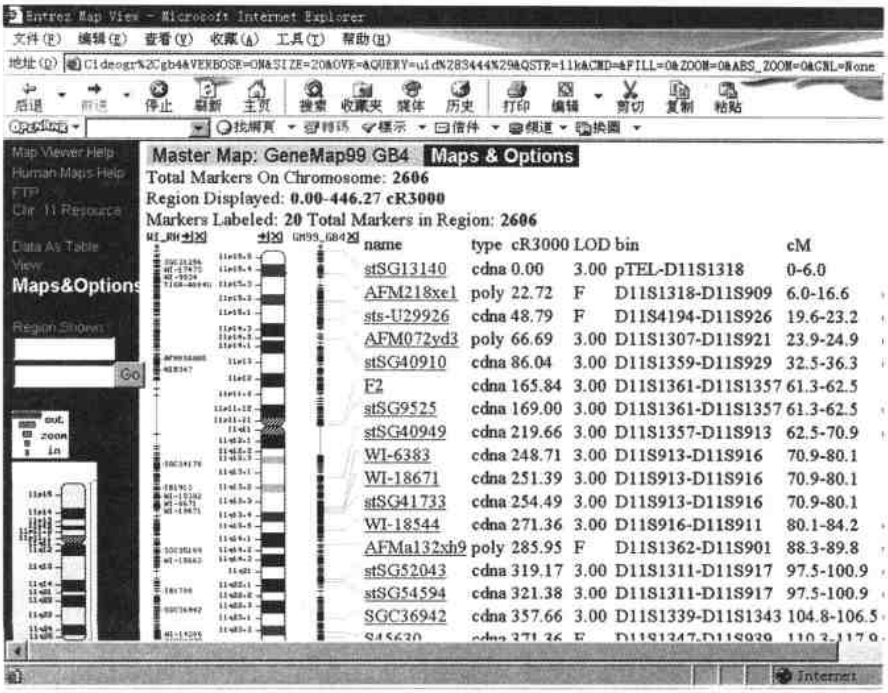
查询后,发现 ILK 含 UniSTS 位标 A002C40。



进一步点击 A002C40,则可以得到有关 ILK 及其染色体定位的详细信息。

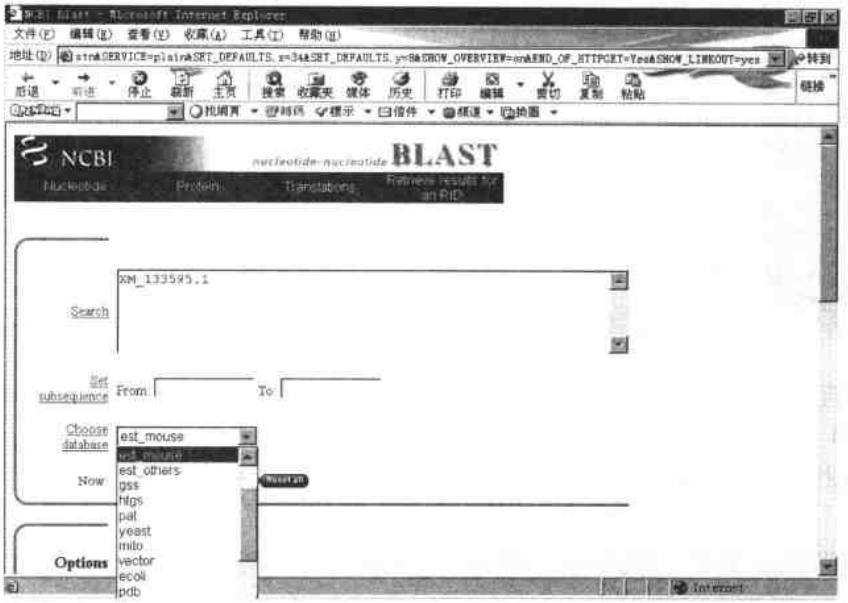


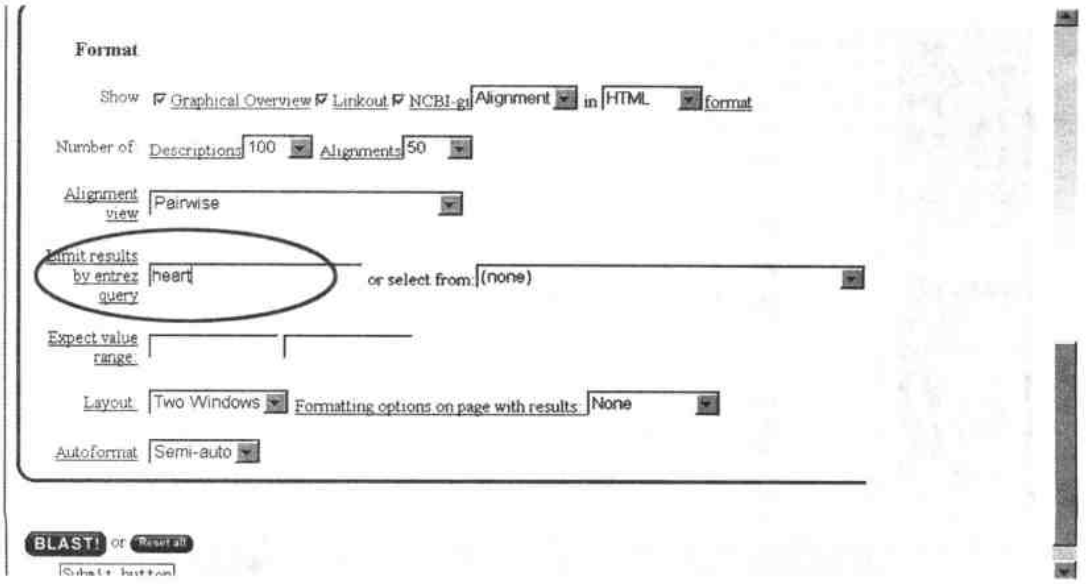
进一步查询,则得到如下定位信息:图中由左到右显示了 Whitehead RH、Ideogram 、GeneMap99 – GB4 三张图。图中红点处代表 ILK 位置,图中 pTEL 表示短臂端粒处,其中 RH 图的位标 D11S1318 代表的含义如下:D 代表 DNA,11 代表染色体号,S 代表单拷贝,1 318 是其位标号码。



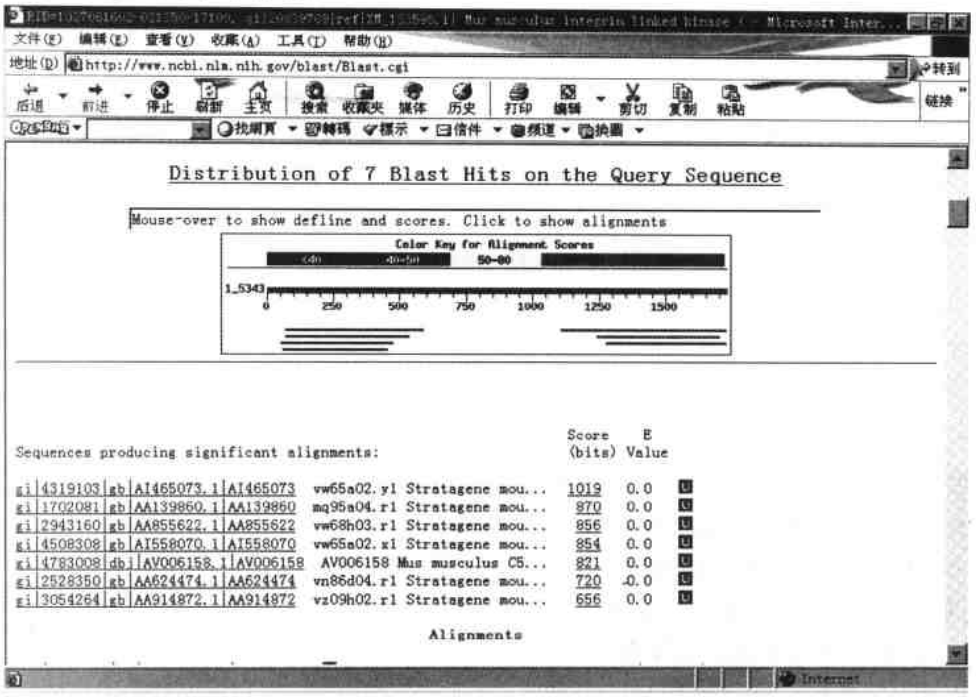
第三节　掌握和熟悉 NCBI 数据库和分析

美国国立生物技术信息中心(NCBI)是一个非常好的网站(欧洲生物信息中心 EBI 和欧洲分子生物学实验室 EMBL 也非常好)。第六章只是非常简单地介绍了 NCBI 数据库运用的一个非常小的方面。实际上 NCBI 包含了文献查询(PubMed)、核酸序列(GenBank)、序列同源性分析(BLAST、PSI－BLAST、PHI－BLAST 等)、疾病数据库(OMIM)、蛋白结构等等各种数据库,其各种数据库通过 Entrez 有机整合在一起,并处于不断更新发展中。熟练结合运用该中心的数据库和软件将大大提高工作效率。例如可将 Entrez 与 BLAST 相结合进行同源分析。如为了解鼠源 ILK 在心脏中的表达,进入 BLAST 主页,选择小鼠 EST 数据库,在 Limit results by entrez query 一栏中输入 heart 并进行 BLAST 检索。





在结果中出现了在小鼠心脏中表达的 EST。当某个基因在各种组织中的表达丰度都比较高时，这种方法将显得非常有利。



（赵　勇）

第七章

核酸及蛋白序列的计算机初步分析

内容提要

生物信息学就是将信息技术运用到生物科学中的一门学科，序列分析是生物信息学分析的重要内容。序列分析包括核酸分析和蛋白序列分析。其中序列的同源/相似性分析、结构域分析、多序列排列和分子进化分析是序列分析的主要内容。最后以某些蛋白激酶为例，说明如何用 Fitch 方法对其进行分析并绘制进化树，将不同的蛋白激酶准确地归为不同家族，以提示它们在进化上的联系。

人类基因组计划产生了海量的数据，这么多数据里面包含了几乎是无穷的有用信息，但是数据不经过分析综合，就不能成为信息，而信息不经过归纳整理就不能成为可利用的知识。如何将人类基因组计划产生的浩瀚数据转变为知识，其中除了运用实验方法外，还必须借助于计算机等信息科学。生物信息学就是随着人类基因组计划应运而生的一门交叉学科。简单地讲，生物信息学就是将信息技术运用到生物科学中的一门学科。在生物信息学分析中，涉及数据库建立、序列分析和发展各种计算方法等，其中最常涉及的就是序列分析。本文就序列分析作一简单介绍。

第一节 概 述

正确理解有关染色体、基因和蛋白质的基本概念是计算机分析的前提。遗憾的是一些在国内期刊发表的相关文章在这些方面有时出现明显的错误，所以有必要简单介绍一下有关概念。

人类基因组应当包括核基因组和线粒体基因组。核基因组即指 22 对常染色体和 1 对性染色体(XX 或 XY)。而属母系遗传的线粒体基因组长度在 16.6 kb。图 7 - 1 中所示为 1 号同源染色体(细胞分裂中期)。每条染色体由两条染色单体组成，每条染色单体含一个 DNA 双螺旋分子。染色体所含的染色质分常染色质和异染色质两种。异染色质主要分布于着丝粒和某些中间部分，一般不含基因或很少有基因，其所含 DNA 主要是一些高度重复序列。常染色质占染色体的绝大部分，基因主要分布于常染色质区。

用显带技术可将染色体的长短臂分别以着丝粒为起点，横向分为若干区带和亚带，如 1p31.1 代表 1 号染色体短臂 3 区 1 带第 1 亚带。Giemsa 染料染色显示的 G 显带(G - bands)与

GC 含量之间有联系。最暗的 G 带处 GC 含量低。另外 GC 含量与重复含量、基因密度有关，GC 含量低的区域，所含基因少。

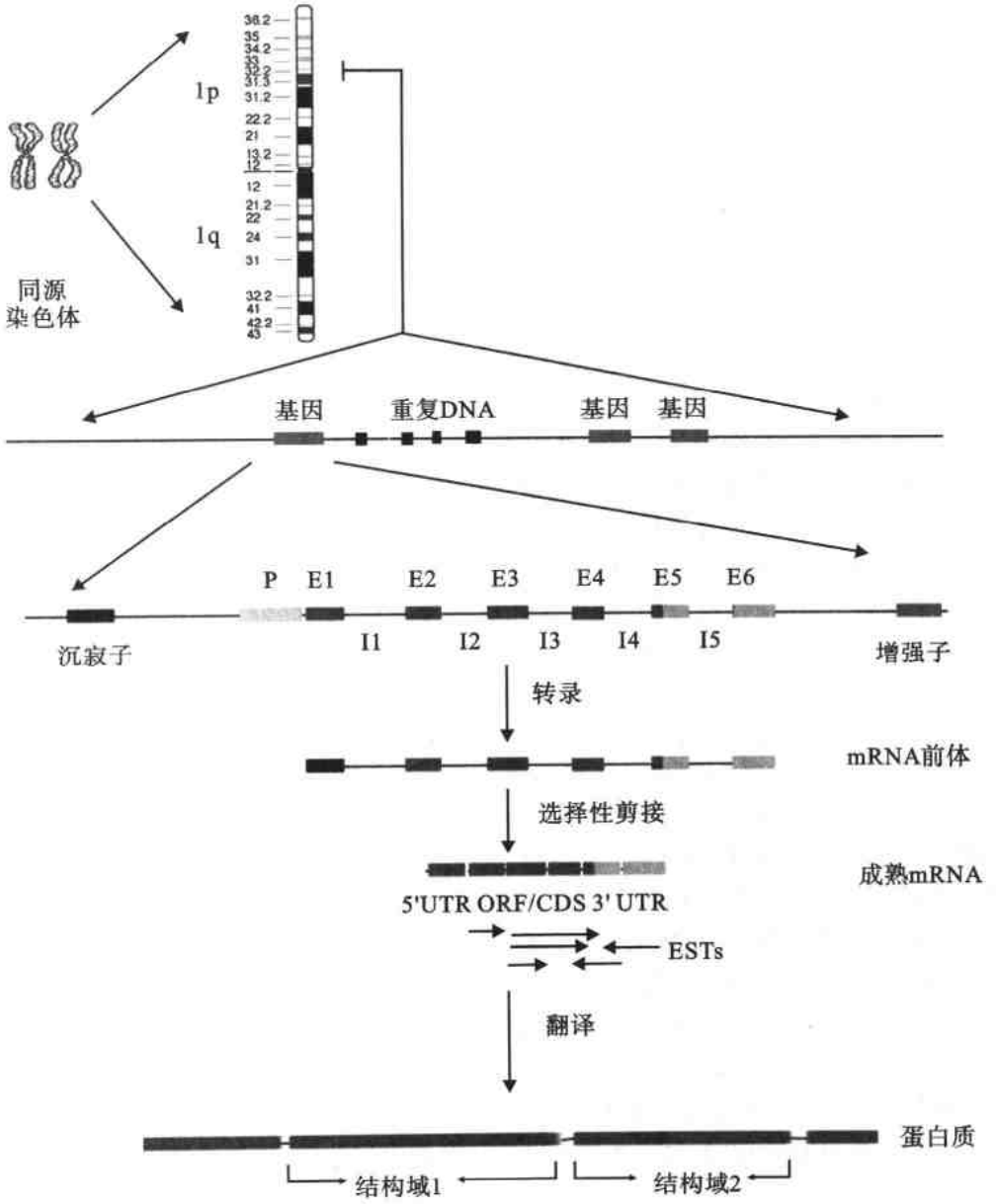


图 7-1　染色体、基因及蛋白示意图

P 代表启动子，E 为外显子，I 为内含子

人类基因组共有 3 200 000 kb(3.2 Gb)。基因在染色体图上的距离单位称为图距，1%重组值(交换值)去掉其百分率的数值定义为一个图距单位，以“厘摩(centimorgan，cM)”表示。整个人类基因组平均遗传长度为 3 200cM(最新估计人类基因组有 3 124.17 Mb(human build 29)，1

cM 约等于 1 000 kb 即 1 Mb。从染色体短臂末端开始向长臂计数。例如定位于 11p15.5 - p15.4的基因 ILK,其基因组序列在 6 028 K ~ 6 046 K 之间,即距短臂末端约 7 cM。染色体在染色体 DNA 上分布有结构基因,调控基因,rRNA 基因,重复序列等。人类基因数若以 30 000 ~ 35 000计算,编码序列长度以 1 400 bp 计(人类编码序列平均长度 1 340 bp),平均的基因组长度约 30 kb,那么编码序列占整个人类基因组序列的 1.5%,而基因组的 1/3 都是能转录的基因序列。除基因序列外,基因组上绝大部分序列为重复序列。

染色体表达结构域(chromosomal expression domains, CED):近年的研究表明染色体存在组织表达基因的区域聚集倾向性。相邻或相近的基因有相关的表达,而且功能相似的基因倾向于出现在染色体中相邻的位置,与原核生物有点类似。

现代的基因概念应当包括基因调控区(启动子和调节序列)。调节序列常散在分布于转录起始位点上下游各约 50 kb 的区域里。启动子通常靠近转录起点,即 DNA 中实际转录合成 RNA 的第一个碱基对,这个位点的核苷酸残基编号为 +1,从转录起点沿 RNA 聚合酶运动的方向称为下游(downstream),核苷酸残基编号依次为 +2, +3 等,而反方向则为上游(upstream),核苷酸残基编号依次为 -1, -2, -3 等。所有的外显子和内含子都转录(但选择性剪接除外),但不是所有的外显子都编码蛋白。mRNA 第一个外显子(最 5′端)至翻译起始位点的序列称为 5′UTR(5′ Untranslated Regions),含核糖体结合位点和翻译调节序列,在某些特殊情况下(如同一基因有多个启动子),含转录调节序列。5′ - UTR 结构决定翻译的效率,有时基因家族 5′ - UTR 的这种多样性可能决定该家族成员的组织及发育阶段特异的调节。复杂的 5′ - UTR(高 G + C 含量和多个 AUG 等)不利于翻译,但可作为不同信号通路分子作用的底物从而提供对重要功能基因(如原癌基因、转录因子基因等)的动力学翻译调节。

CpG 岛(p 仅代表 C 与 G 之间的磷酸根):目前广为接受的定义最初由 Gardiner - Garden 等提出。CpG 岛是一段约 200 bp 长的 DNA,其 G + C 含量大于 50%,并且观察到的 CpG 与预期 CpG 比值(ObsCpG/ExpCpG)不小于 0.6,它存在于几乎所有看家基因和约 40%组织特异性基因的启动子和外显子区域。CpG 岛在基因沉寂、基因组印迹、X 染色体灭活、肿瘤发生等具有重要作用。但是由于有些重复序列如 Alu 序列等也含与 CpG 岛相似序列,而这些与 CpG 岛相似的序列很可能不具有与真正 CpG 岛同样的功能,所以寻找核酸的 CpG 岛时必须排除重复序列。据估计整个基因组共有 28 890 个 CpG 岛,大部分染色体含 5 ~ 15 个 CpG 岛/Mb,平均有 10.5 个 CpG 岛/Mb。欧洲生物信息中心(EBI)有专门的程序 CpGPlot/CpGReport/Isochore (http://www.ebi.ac.uk/emboss/cpgplot/)来识别 CpG 岛。

cDNA 全长:由于建库的原因,很多情况下获得的 cDNA 并非全长,只是有完整的读码框,习惯上称的 cDNA 全长是指有完整编码框,而且其 cDNA 长度与 Northern Blot 结果基本一致。

开放读码框(CDS/ORF):指由蛋白质合成的起始密码开始,到终止密码子为止的一个连续编码序列。

蛋白结构域及外显子:见本章下述。

同源性(homology)、相似性(similarity)、一致性(identity)和保守性(conservation):同源性是指由于有共同祖先所以具有的相似性。序列间有无共同祖先是有或无的关系,因此序列间同源性也是有或无,而不应出现"70%同源"、有"很高同源性"这些错误说法。相似性是指核酸或蛋白序列相关的程度,这种程度用一致性和/或保守性的程度来衡量。很显然序列间相似程度可以不一,可以说两个序列"70%相似"。另外需要考虑的是,具有显著相似性的序列是整个序列间或序列内部的保守结构域,是用哪种计算方法(程序)找到的。有显著相似性的蛋白在绝大

多数情况下是同源的。一致性是指核酸或蛋白序列不变的程度。例如两个序列有 40% 的一致性。保守性是指在氨基酸序列的某一个特定位点保留有原始残基的物理化学性质的现象，该概念很少用于 DNA。

第二节　核酸序列分析

一、数据库同源性或相似性搜索

最常用的序列同源性或相似性查询工具是 BLAST（http://www.ncbi.nlm.nih.gov/BLAST/）和 FASTA（http://www.ebi.ac.uk/fasta33/）。

BLAST（Basic Local Alignment Search Tool）即基本的局部对准搜索工具。尽管随着数据库数据的指数增长，BLAST 检索速度似乎较前几年略慢，但总体来说其搜索速度还是非常快。不同长度的核酸序列及不同的检索目的所选用的 BLAST 程序不同。下表简要列出了各种不同情况下选用不同 BLAST 程序的一些要点。

表 7-1　核酸序列的 Blast 分析

长　度	数据库	目　的	BLAST 程序
不小于 20 bp	核酸	证实序列	MEGABLAST 可接受批量查询标准的 BLAST（blastn）
		寻找相似序列	标准的 BLAST（blastn）
		在翻译数据库中寻找与翻译的被检序列相似的蛋白序列	翻译的 BLAST（tblastx）
	蛋白	在蛋白数据库中寻找与翻译的被检序列相似的蛋白序列	翻译的 BLAST（blastx）
7～20 bp	核酸	寻找引物结合位点或短的寡核苷酸基序	寻找短几近匹配序列

二、读码框分析

拿到一个核酸序列，为了确证其是否为蛋白编码基因，可首先按六种阅读框将其翻译成蛋白质，看其读码框的长短。NCBI 的 ORF finder 是一个很方便的寻找核酸读码框的工具。

尽管有短至几十个氨基酸的分泌肽（已切除其信号肽序列，例如人的 alpha－ANP 仅 28 aa），也有极少数几十个氨基酸的非分泌型短肽（近年 PNAS 报道过），但就笔者经验而言，一般来说如果阅读框短至 300 nt 以下，并且核酸序列同源性分析在其他物种没有明显同源序列的话，那么该核酸序列不大可能是一个真正的蛋白编码基因。

这里需提一下的是假基因。在多基因家族中（即有多个旁系同源物），某些成员不能转录或不能翻译，所以不产生有功能的基因产物，其 DNA 序列与有功能的基因相似，这样的成员称为假基因。一般认为假基因由于逆转录酶介导的转座作用产生。假基因一般缺乏完整的读码

框(在一个较长读码框中有终止密码)。

三、寻找启动子区及其核酸序列中的功能位点

启动子通常位于转录起始点上游 200 bp 左右(有许多计算机软件可以预测)。基因的时空特异性分布在一定程度上由不同的反式作用因子和顺式调控元件协同决定。通过计算机寻找这些反式作用因子的结合位点和/或保守序列元件可能为基因的功能或分子通路研究带来线索。著名期刊 *Nature*、*Science* 等已有许多这方面成功的报道。但由于转录因子结合位点有时可落在前几个内含子区(例如 *P*53 可以与生长抑制和 DNA 损伤诱导基因 GADD45 的第 3 个内含子区结合),这就增加了分析的复杂性。

四、基因的染色体定位

基因定位的有关方法见前述(第六章第二节)。知道基因的染色体定位,应分析其是否落在染色体的疾病候选区、该基因的外显子和内含子数目、重要蛋白结构域所对应的外显子有无多态性、比较基因组学研究、发现保守的调控区域等,可方便实验设计,例如辅助基因打靶策略的设计等。

第三节 蛋白序列分析

一、同源性或相似性分析

同核酸序列分析类似,在进行蛋白的序列分析时,首先进行数据库同源性分析,可以得到一个大概的印象,即被检序列大概与哪些蛋白家族相关,是否与已知序列相同。同样,同源性分析也须借助 BLAST 工具(或 FASTA),下表列出了运用 BLAST 的不同情况:

表 7-2 氨基酸序列的 BLAST 分析

长度	数据库	目的	BLAST 程序
不小于 15 个氨基酸	蛋白数据库	证实或寻找与被检序列相似的蛋白序列	标准的蛋白 BLAST (blastp)
		寻找蛋白家族成员或建立自定的位置特异的分数矩阵	PSI - BLAST
		寻找含给定蛋白序列模式的相似蛋白序列	PHI - BLAST
	保守结构域数据库	寻找被检序列的保守结构域	CD - search (RPS - BLAST)
		寻找被检序列的保守结构域并证实含相似结构域组成的蛋白	保守结构域组成截取工具(CDART)
	核酸	在翻译的核酸数据库中寻找相似蛋白	翻译的 BLAST (tblastn)
5~15 氨基酸	蛋白	寻找肽基序	寻找短几近匹配序列

二、氨基酸序列的一般特性分析

在进行蛋白序列分析时，瑞士蛋白专家分析系统（ExPASy）是一个非常好的网络工具（http://www.expasy.ch/tools/），其几乎囊括了所有的用于蛋白分析的计算工具。

分析一个新基因的蛋白产物，如等电点 pI 和蛋白分子量，氨基酸组成等。在蛋白的一般特性分析中，SAPS（蛋白序列的统计分析，Statistics on protein sequences，欧洲生物信息学研究中心的 http://www.ebi.ac.uk/saps/或 http://www.isrec.isb-sib.ch/software/SAPS-form.html）是一个比较全面的蛋白一般特性分析软件。它可以对提交的序列给出大量全面的分析数据，包括氨基酸组成统计、电荷分布分析、电荷聚集区域、高度疏水区域、跨膜区段等。读者可根据研究目的进行选择。

三、蛋白基序和结构域分析

蛋白结构域（domain）是指蛋白质结构中具有独立三级结构的部分，并具有特定的功能。在许多情况下，一个蛋白保守的（亚）结构域大致对应于一个（或数个）外显子。而蛋白基序（motif）是结构域的亚单元，因而序列较结构域短，它体现结构域的各种生物学作用。通常情况下蛋白基序要较蛋白结构域更为保守。在某些文献中，这两个概念常被混用。例如：所有蛋白激酶都有 250～300 氨基酸的激酶结构域，在空间结构域上，几乎所有蛋白激酶结构域都分一大一小两叶，有相同的空间结构。功能上蛋白激酶可催化磷酸基团的转移。蛋白激酶结构域又分 11 个亚结构域，其中亚结构域 VII 的 DFG 三联体在所有的蛋白激酶中均存在，可以称为蛋白基序，该基序参与 ATP 结合。

蛋白基序和结构域分析是蛋白序列分析中极其重要的方面。蛋白结构域常作为蛋白功能单位决定了该蛋白主要的生物学功能。不同功能的蛋白家族具有其独特的蛋白结构域。各种不同的结构域各司其职，即如果一个蛋白有多种功能，其所含的多个结构域分别执行不同的功能。例如酪氨酸蛋白激酶 Src 含 SH3、SH2 和激酶结构域，激酶结构域参与磷酸化反应，而 SH3、SH2 参与蛋白的相互作用。

有些结构域在个体发育或疾病中具有重要功能，因而也使含这些结构域的蛋白具有同样的重要作用。本书附录有蛋白分析的权威数据库。

四、特殊结构或结构特征分析

此处所指的特殊结构是指有无信号肽、卷曲螺旋（coiled coil）、跨膜结构及偏性组成区等。

1. 卷曲螺旋是由 2 或多个 α-螺旋组成的寡聚化基序，可以左手螺旋的方式相互缠绕。卷曲螺旋常常与其他卷曲螺旋相互作用，而且卷曲螺旋通常不需要其他蛋白结构域介导即能与其他蛋白发生相互作用。卷曲螺旋是一自主折叠单位，能独立折叠。许多蛋白，包括信号分子（如激酶等）、转录因子和结构蛋白等都含卷曲螺旋。卷曲螺旋的预测方法是将待检测序列与已知的平行双链卷曲螺旋数据库进行比较，得到相似性得分后，据此算出序列形成卷曲螺旋的概率（COILS 的网址是：http://www.ch.embnet.org/software/COILS-form.html）。

2. 跨膜结构预测的基本原理是跨膜蛋白具有一或多个疏水区域。TMpred 可预测蛋白质的跨膜区域和在膜上的取向（TMpred 的网址是：http://www.ch.embnet.org/software/TMPRED-form.html）。

3. 除很少的蛋白外，绝大多数的蛋白都在胞浆核糖体中合成，这些蛋白随后的命运是分

泌到胞外还是留在胞浆或入核等则取决于蛋白序列中的分拣信号。蛋白在细胞内的不同细胞器区域转移有两种截然不同的机制：一种是蛋白在膜蛋白转移器的帮助下，改变其拓扑学结构，打开折叠穿过膜，如部分蛋白从胞浆转移到 ER 腔即是一个例子。另外一种机制是蛋白通过运输颗粒运输，例如可溶性蛋白从 ER 腔到高尔基氏器，在此种情况下蛋白没有跨膜过程。这两种过程均由蛋白本身的分拣信号控制。目前认为蛋白质有两种分拣信号，一种即是广义的信号肽，另外一种是信号补丁。信号肽一般是一段连续的长约 5～60 aa 的序列，常常在蛋白分拣后被切除。例如有 N 端信号肽的胞外分泌蛋白，在其分泌到胞外的过程中，其 N 端信号肽被切除。而构成信号补丁的氨基酸序列在线性氨基酸序列上距离较远，只是空间上接近。目前计算机预测的都是信号肽而非信号补丁，后者分析预测非常困难。Albert 将核定位信号（nuclear localization signal）也归为一种信号肽，许多核蛋白含多个核定位信号，而且核定位信号在核蛋白入核后并不被切除，这可能是由于核蛋白在每个细胞分裂周期后要反复出入核。另外值得注意的是，不是所有的入核蛋白都必须要有核定位信号，这与细胞、蛋白当时所处的具体环境有关。例如 MAPK 不含核定位信号，但其可以在激活时由胞浆入核，激活转录因子从而发挥其功能。而 NLK 激酶也不含核定位信号，其在不激活的情况下大部分也位于核内。

SignalP：预测蛋白质序列中信号肽的剪切位点（SignalP 的网址是：http://www.cbs.dtu.dk/services/SignalP/）。

五、直系同源物（Ortholog）和/或旁系同源物（Paralog）分析

直系同源物是指在物种分化过程中源于同一共同祖先基因的不同物种的同源序列，如鼠与人的 β－globin 可称为直系同源物。旁系同源物是指在同一物种内由于基因复制而得到的同源序列，如人的 β 与 γ－globin。显然旁系同源物之间也是同源的。

进行数据库同源性分析后，判断基因有无直系同源物和/或旁系同源物，可以提供功能线索。一般情况下，直系同源物具有相同或相似的功能，而旁系同源物的功能则不同。然而直系同源物在不同物种中的时空特异性表达也可能有异，所以某些情况下不同物种的直系同源物的功能也可能不同。

在实际操作中，判断两个来源于不同物种的基因是否为直系同源物，大致有以下标准：

1. 在两个物种间交互比较时，它们有最高的配对一致性（即序列 A 在与另一物种同源比较时，与序列 B 的相似性最高；反之亦然。）。

2. 二者相似性的 E 值（即假阳性的可能性）应小于 0.01。

3. 这种相似性至少横跨其中一个基因序列的 60%，因为有基因融合或基因分裂事件的存在，但不一定要覆盖两个基因的大部分区域。

若两个源于不同物种的基因能满足这些条件，那么它们很可能为直系同源物。NCBI 的 COGnitor 可以识别直系同源物（也包括旁系同源物）（http://www.ncbi.nlm.nih.gov/COG/）。

六、二级结构的预测

蛋白的功能是由其空间结构决定的，即具有相同或相似空间结构的蛋白，不管其序列相似性如何，都具有相同或相近的功能。目前尽管有通过蛋白质二级结构来预测蛋白质功能的尝试，但是这种方法尚不成熟，仅可将对蛋白质的二级结构预测用于辅助抗体的制备。而随着越来越多的对蛋白质空间结构的测定和三级结构预测方法的改进，未来通过结构的分析进行功能预测将很有前景。

欧洲分子生物学实验室的 PredictProtein 服务器是一个比较好的二级结构预测软件。

第四节　多重序列排列和系统发育分析

一、多重序列排列

简单地说，多重序列排列就是将两个或两个以上的序列对准，获得最大水平的一致性（在氨基酸，称保守性），以评价序列相似性的程度和同源性的可能性。因此，多重序列对比不限于那些有进化关系的序列，序列间只要相关就可以，这与进化分析不同。相关蛋白可以指同源（homology）或趋同（convergence）。在这两种情况下都可以用序列对排。同源蛋白质有共同祖先，趋同进化的蛋白质是独立进化，趋于有共同的序列特征或空间结构，所以有相同的功能。趋同进化有下列两个例子：各种切除酯和多肽链的水解酶，它们具有共同的催化三联体结构；古细菌、真细菌和真核蛋白的 DNA 结合结构域都具有螺旋－转折－螺旋（HLH）结构。

多序列排列有两个目的：一是发现序列间保守的结构，二是在有些情况下可推断序列间的进化关系。

多序列排列分全局排列（global alignment）和局部排列（local alignment）两种。目前使用最广泛的多序列比对程序是 Clustal W（Clustal X 是其 Windows 版本）。Clustal W 是一种渐进的比对方法，在序列两两比对的基础上构建距离矩阵，根据距离矩阵由 NJ 方法构建进化指导树，并对序列进行加权。之后从最相似的一对序列开始，逐步引入临近的相似序列并不断重新构建比对，直到完成全部序列的比对。著名分子生物学分析软件包 GCG 的 pileup 程序与 Clustal W 相似，均属全局排列。

有时序列间仅有一个相似的区域，其他区域完全不同，如酶的活性位点或转录因子的 DNA 结合位点。此时就需要进行局部排列，这种排列可避免插空排列。由 NCBI 开发的 MACAW 软件就是局部排列软件。

二、分子进化分析

现代科学文献中，运用分子进化进行基因和/或蛋白的功能推断越来越普遍，这是因为需要使生物学更具有可预见性。如同元素周期表可使化学家预言化学反应一样，进化分析可以使生物学家对基因、蛋白或生物的性状、行为做出预测。分子进化是通过核酸和/或蛋白的序列差异来推导序列间的祖先进化相互关系的。显然系统发育分析的前提是所有的序列必须有一共同祖先，这样的推断才有意义。

由分子进化推断出的进化树有两种，一种是有根树，在这种树形结构中，祖先序列以树根（或树干）的形式出现，并有明确的序列进化过程。另一种是无根树，它仅仅显示树中序列的相互进化关系，并不推断祖先序列的位置或进化途径，因此它不能给出进化过程。

（一）分子进化分析的序列要求

1. 序列是同源的，即所有序列均起源于同一祖先序列。例如由于所有的蛋白激酶都起源于同一共同祖先，所以不同的蛋白激酶间可以进行进化分析。

2. 序列比对中，每一个位点都是同源的。例如对蛋白激酶的氨基酸序列进行排列时，所有序列的 ATP 结合位点要相对应。

3. 在接受分析的序列组中，每一个序列组的系统发育史是相同的。不能将核序列与线粒体序列进行比较。

4. 样品数足够大，至少 6 个以上，样品数太少不足以给出有用信息。

5. 样品序列之间的差异合适，此差异既不太大，也不太小。如序列相互间刚有进化差异，但还没有改变太多，这些序列可能在横跨整个序列全长的地方都很相似，这样就不能将那些有重要功能的保守区域与那些未有足够时间差异的区域区分开。如序列间差异太大（甚至可能不相关），几乎没有相似区域，或序列间也可能有相关区域，但从这些差异太大的序列中不能发现这些区域，除非有更多序列或有更敏感的排列方法。

6. 样品序列之间的差异包含了所研究的系统发育信号。

另外不同的方法有不同的假设。同时对样品序列是否随机进化、每个位点进化是否独立等问题也需要考虑。

（二）分子进化分析的用途

分子进化分析可用于：

1. 基因或蛋白质的功能推断。

2. 蛋白序列的保守结构域分析。

3. 种系间保守的调控元件（如基因的启动子区的调控序列）的发现。

（三）分子进化分析的大致过程

由于分子进化非常复杂，对于一般的科研工作者，很难做到对其来龙去脉非常清楚地了解。在分析自己工作中的核酸或蛋白序列时，参考以前权威杂志已发表的相关序列所用的分析方法是一个比较好的办法。这里仅给出一些比较基本的概念，有兴趣的读者可参考有关的专业书籍。

1. 多重序列排列比对。有关原理见前述，但需要注意的是不同相似程度的序列比较时，须采用不同的计分矩阵。序列排列比对的质量决定了后续建树的可信性。另外发表文章时须说明采用什么程序、有无手工排列等。

2. 决定建树方法。

（1）构建进化树的算法主要分为两类：即基于距离和基于特征符的方法。

基于距离的方法计算序列间的差异数量，并将其看作进化距离，通过某种聚类算法，从最相似（距离最短）的一对序列开始，经距离矩阵推导出进化树，或者通过将总的树枝长度最小化而优化出进化树。进化树的拓扑结构则由序列间的进化距离决定，进化树枝的长度代表着进化距离。基于距离的方法包括不加权配对算术方法（UPGMAM）、相邻连接方法（NJ）、Fitch－Margoliash（FM）和最小进化方法（ME）。

基于特征符的方法是指进化树的拓扑结构由序列中的每个碱基/氨基酸的状态决定（例如一个核酸序列上可能包含多个酶切位点，而每个酶切位点的存在与否是由几个碱基的状态决定的）。基于特征符的方法包括最大节约方法（MP）和最大似然方法（ML）。

（2）各种方法大致的适用条件和优缺点：总体来说，基于距离的方法适于免疫学数据、频率数据或有些方法不明的数据。计算很快，易于用统计学检验（如 Bootstrapping）。

UPGMAM 方法首先计算所有序列间的距离（差异分数）矩阵，随后由最小距离（差异分数）

的两对序列以有树根的形式被连在一起，该序列对的每一个序列与第三个节点距离的平均值被用于下一轮距离矩阵的重复计算。该过程反复重复直至所有序列都被放到树枝上。所以该方法总能得到一棵有根的树。显然该方法有一非常关键的假设即在进化过程中所有核苷酸/氨基酸都有相同的变异率，也就是存在着一个分子钟，所有序列都以同样速率突变。这种算法得到的进化树相对来说不是很准确，现在已经很少使用。

相邻连接方法是基于距离方法的一个变异，可创建一棵无根的树。同样它首先计算所有序列间的距离矩阵，有最小距离（差异分数）的两对序列以有树根的形式被连在一起，随后这样的序列对被当做一个序列处理，再重新进行距离计算，该过程反复重复，直至所有序列都被放到树上。该方法也有分子钟的假设即所有序列都以同样速率突变。其计算快捷，但缺点是序列上的所有位点都被同等对待，而且所分析的序列的进化距离不能太大。

Fitch 和 Margoliash 法寻找符合某种标准的进化树，而不是寻找最相似的序列对，产生无根树。

最大节约法是一种在重建祖先关系方面很流行的算法。最大节约方法考察数据组中序列的多重比对结果，它创建的进化树试图通过最少的进化改变来解释观察到的数据，例如在某一特定位点有 15 次突变，有可能在第一和最后一次的突变使得在该位点的碱基回复到与原来一样的碱基。最大节约法假设在该位点仅发生一次突变（一步）。它适用于符合以下条件的多序列：①所要比较的序列的碱基差别小；②对于序列上的每一个碱基有近似相等的变异率；③没有过多的颠换/转换的倾向；④被检验序列的碱基数目较多（大于几千个碱基）。然而问题是许多最大节约进化树可能不是惟一的，如果树中有许多步骤，很难进行统计学检验，而且此种方法计算比较耗时。

最大似然方法分析序列则不需以上的诸多条件，用该方法推导出的进化树能够以最大的概率考察多重比对结果，或者说该方法创建所有可能的进化树，随后用统计学方法进行评价哪一棵树是最有可能的。但此种方法计算比较耗时。

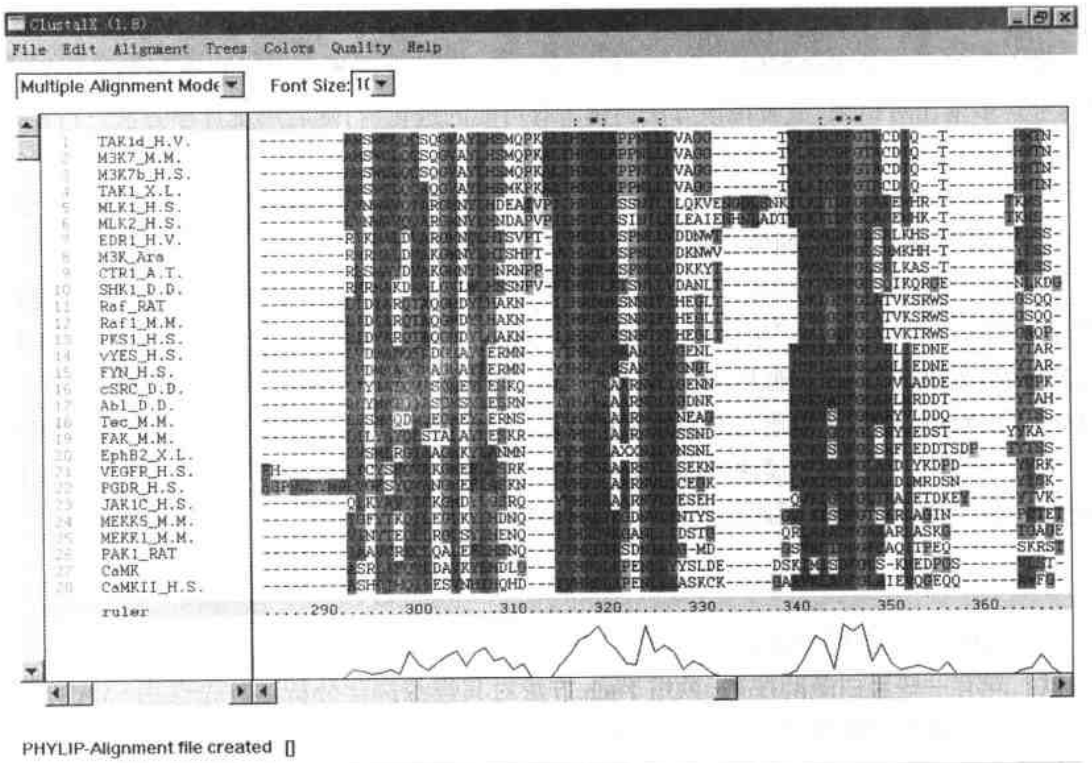
3.确定树根和评估进化树。上述的建树方法所产生的都是无根树（进化树没有进化的极性）。为了评估进化假说，通常必须要确定进化树的树根。

评估进化树可以有自展法（bootstrap）和折刀法（Jackknife method）。而自展法在分子进化的检验中应用较广。自展法用来估计推导出的进化树的可能性。原则上，被比较的序列在有些地方被随机处理，它们的相似性用标准计分矩阵再次比较。涉及随机取代计数 N 个字符从而创建一新的数据集，该数据集与原始数据的大小相同，但是有些字符不考虑，有些则被两次处理。对每个序列，大部分自展法都随机处理 100 次，并用随机处理的统计学与原始序列比较分值进行比较。所得结果是对原始序列比较分值随机发生的可能性的估计。

三、举　例

下面给出一个进化分析的简单例子，目的是说明大概的操作过程。有兴趣的读者可以自行一试。现有一些蛋白激酶序列，欲用 Fitch 方法对其进行进化分析。首先选用 SMART 程序将各个蛋白激酶的激酶结构域的边界找出。序列存成 .txt 文件；随后用 CLUSTAL X 将其打开，并进行多序列排列，同时输出 .phy 文件以备用 PHYLIP 软件包进行进化分析。

domain - 记事本
文件(F) 编辑(E) 查看(V) 帮助(H)
>PKA_H.S.
ERIKTLGTGSFGRVMLVKHKETGNHYAMKILDKQKVVKLKQIEHTLNEKRILQAVNFPFLVKLEFSFKDNSNLYMVMEYVPGGEMFSHLRRIGRFSEPH
ARFYAAQIVLTFEYLHSLDLIYRDLKPENLLIDQQGYIQVTDFGFAKRVKGRTWTLCGTPEYLAPEIILSKGYNKAVDWWALGVLIYEMAAGYPPFFAD
QPIQIYEKIVSGKVRFPSHFSSDLKDLLRNLLQVDLTKRFGNLKNGVNDIKNHKWF
>PKC_H.S.
NFLMVLGKGSFGKVMLADRKGTEELYAIKILKKDVVIQDDDVECTMVEKRVLALLDKPPFLTQLHSCFQTVDRLYFVMEYVNGGDLMYHIQQVGKFKEP
QAVFYAAEISIGLFFLHKRGIIYRDLKLDNVMLDSEGHIKIADFGMCKEHMMDGVTTRTFCGTPDYIAPEIIAYQPYGKSVDWWAYGVLLYEMLAGQPP
FDGEDEDELFQSIMEHNVSYPKSLSKEAVSICKGLMTKHPAKRLGCGPEGERDVREHA
>CaMKII_H.S.
QLFEELGKGAFSVVRRCVKKTSTQEYAAKIINTKKLSARDHQKLEREARICRLLKHPNIVRLHDSISEEGFHYLVFDLVTGGELFEDIVAREYYSEADA
SHCIHQILESVNHIHQHDIVHRDLKPENLLLASKCKGAAVKLADFGLAIEVQGEQQAWFGFAGTPGYLSPEVLRKDPYGKPVDIWACGVILYILLVGYP
PFWDEDQHKLYQQIKAGAYDFPSPEWDTVTPEAKNLINQMLTINPAKRITADQALKHPWV
>CaMK I_H.S.
DFRDVLGTGAFSEVILAEDKRTQKLVAIKCIAKEALEGKEGSMENEIAVLHKIKHPNIVALDDIYESGGHLYLIMQLVSGGELFDRIVEKGFYTERDAS
RLIFQVLDAVKYLHDLGIVHRDLKPENLLYYSLDEDSKIMISDFGLSKMEDPGSVLSTACGTPGYVAPEVLAQKPYSKAVDCWSIGVIAYILLCGYPPF
YDENDAKLFEQILKAEYEFDSPYWDDISDSAKDFIRHLMEKDPEKRFTCEQALQHPWI
>MLCK_H.S.
DIEERLGSGKFGQVFRLVEKKTRKVWAGKFFKAYSAKEKENIRQEISIMNCLHHPKLVQCVDAFEEKANIVMVLEIVSGGELFERIIDEDFELTERECI
KVMRQISEGVEYIHKQGIVHLDLKPENIMCVNKTGTRIKLIDFGLARRLENAGSLKVLFGTPEFVAPEVINYEPIGYATDMWSIGVICYILVSGLSPFM
GDNDNETLANVTSATWDFDDEAFDEISDDAKDFISNLLKKDMKNRLDCTQCLQHPWL
>CDK1_H.S.
TKIEKIGEGTYGVVYKGRHKTTGQVVAMKKIRLESEEEGVPSTAIREISLLKELRHPNIVSLQDVLMQDSRLYLIFEFLSMDLKKYLDSIPPGQYMDSS
LVKSYLYQILQGIVFCHSRRVLHRDLKPQNLLIDDKGTIKLADFGLARAFGIPIRVYTHEVVTLWYRSPEVLLGSARYSTPVDIWSIGTIFAELATKKP
LFHGDSEIDQLFRIFRALGTPNNEVWPEVESLQDYKNTFPKWKPGSLASHVKNLDENGLDLLSKMLIYDPAKRISGKMALNHPYFNDLDNQI
>CDK6_H.S.
ECVAEIGEGAYGKVFKARDLKNGGRFVALKRVRVQTGEEGMPLSTIREVAVLRHLETFEHPNVVRLFDVCTVSRTDRETKLTLVFEHVDQDLTTYLDKV
PEPGVPTETIKDMMFQLLRGLDFLHSHRVVHRDLKPQNILVTSSGQIKLADFGLARIYSFQMALTSVVVTLWYRAPEVLLQSSYATPVDLWSVGCIFAE
MFRRKPLFRGSSDVDQLGKILDVIGLPGEEDWPRDVALPRQAFHSKSAQPIEKFVTDIDELGKDLLLKCLTFNPAKRISAYSALSHP
>ERK1_RAT
TQLQYIGEGAYGMVSSAYDHVRKTRVAIKKISPFEHQTYCQRTLREIQILLRFRHENVIGIRDILRAPTLEAMRDVYIVQDLMETDLYKLLKSQQLSND
HICYFLYQILRGLKYIHSANVLHRDLKPSNLLINTTCDLKICDFGLARIADPEHDHTGFLTEYVATRWYRAPEIMLNSKGYTKSIDIWSVGCILAEMLS
NRPIFPGKHYLDQLNHILGILGSPSQEDLNCIINMKARNYLQSLPSKTKVAWAKLFPKSDSKALDLLDRMLTFNPNKRITVEEALAHPYL
>ERK3_RAT
MDLKPLGCGGNGLVFSAVDNDCDKRVAIKKIVLTDPQSVKHALREIKIIRRLDHDNIVKVFEILGPSGSQLTDDVGSLTELNSVYIVQEYMETDLANVL


ClustalX (1.8)
File Edit Alignment Trees Colors Quality Help
Multiple Alignment Mode
Font Size:
TAK1d_H.V.
M3K7_M.M.
M3K7b_H.S.
TAK1_X.L.
MLK1_H.S.
MLK2_H.S.
EDR1_H.V.
M3K_Ara
CTR1_A.T.
SHK1_D.D.
Raf_RAT
Raf1_M.M.
PKS1_H.S.
vYES_H.S.
FYN_H.S.
cSRC_D.D.
Abl_D.D.
Tec_M.M.
FAK_M.M.
EphB2_X.L.
VEGFR_H.S.
PGDR_H.S.
JAK1C_H.S.
MEKK5_M.M.
MEKK1_M.M.
PAK1_RAT
CaMK
CaMKII_H.S.
ruler
.....290.....300.....310.....320.....330.....340.....350.....360.....
PHYLIP-Alignment file created []

(彩图见封三)

PHYLIP 的功能强大，主要包括 5 个功能程序组：

1. 分子序列数据的分析程序。

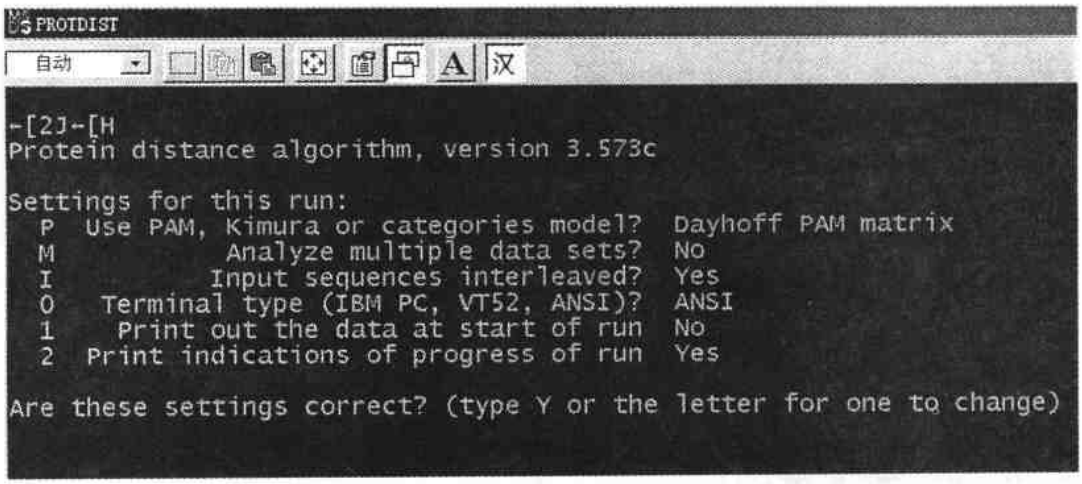
2. 序列数据转变成距离数据后，对距离数据分析的程序（包括 Fitch、Kitch 和 NJ 方法）。

3. 对基因频率和连续字符分析的程序。

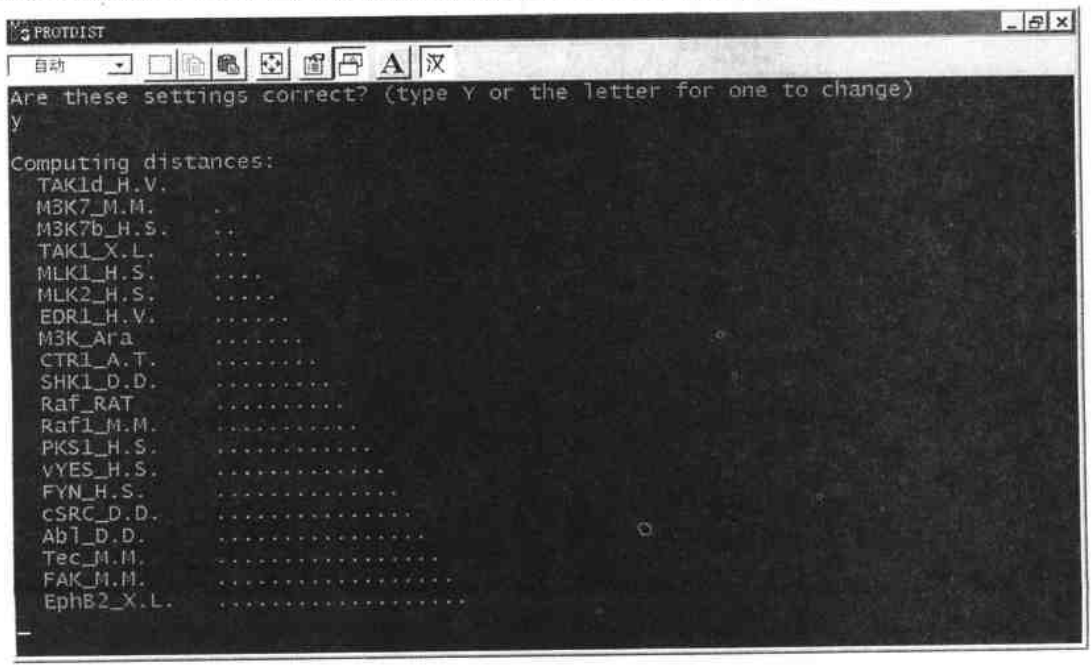
4. 把序列的每个碱基/氨基酸独立看待（即只有 0 和 1 的状态）时，对序列进行分析的程序。

5. 绘制和修改进化树的软件。

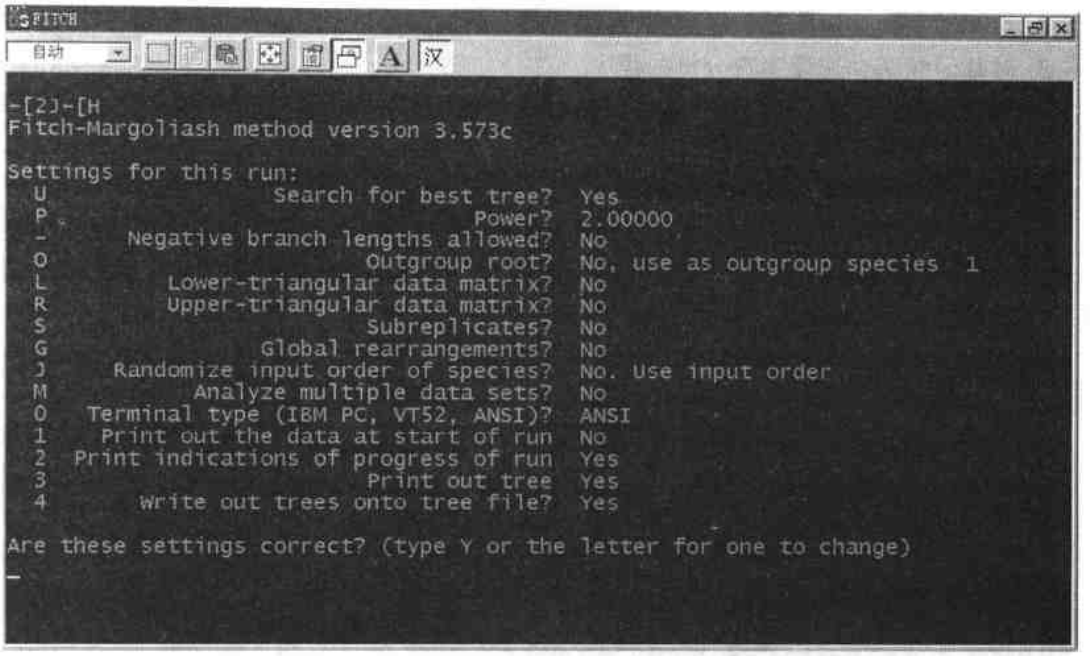
首先将上面排列好的带有后缀 .phy 的序列文件放到安装有程序 PHYLIP 目录下，并更名为 infile。先打开 Protdist 程序以便进行蛋白矩阵计算。打开 Protdist 程序时出现这样一对话框：



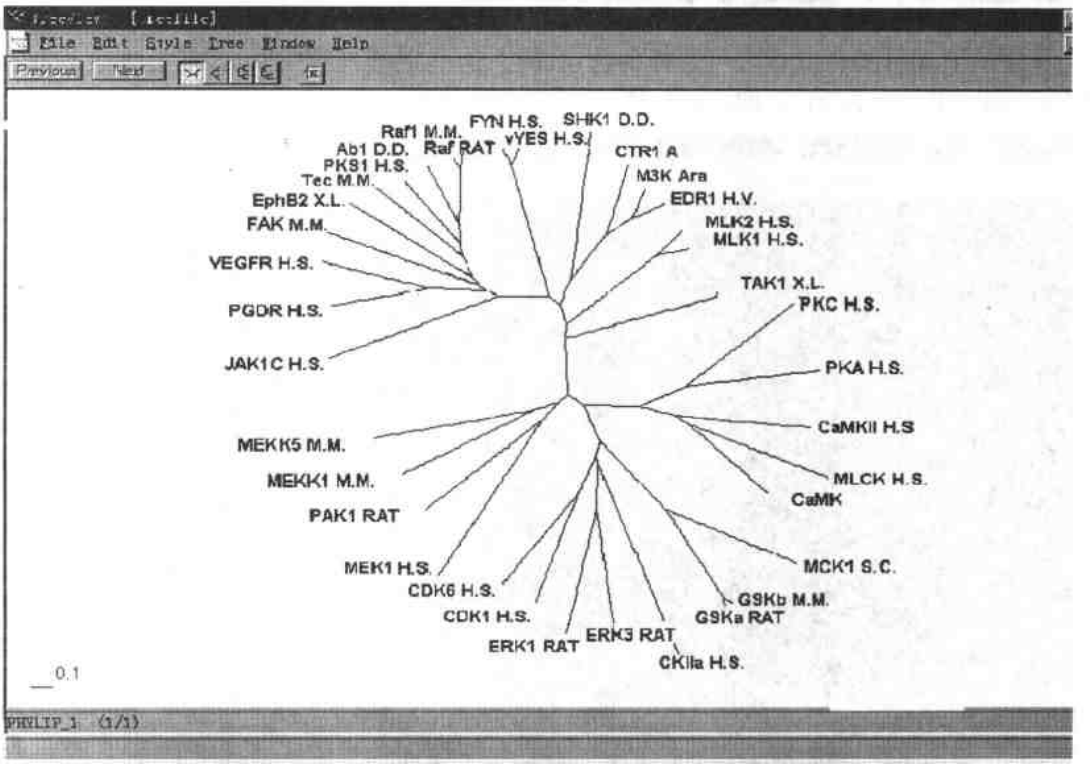
图中的 P、M、I、O、1、2 代表可选择的选项，键入这些字母，程序的条件就会发生改变。选择默认参数，键入 Y 按回车，即开始计算，计算结果被计算机自动命名为文件 outfile。



当我们获得了蛋白矩阵后，就可以用距离的方法构建进化树。这里我们采用 Fitch 的方法。将上述蛋白矩阵文件更名为 infile，随后打开 Fitch 程序，出现如下对话框：



选默认参数，即开始进行建树过程。所得建树文件被计算机自动命名为 treefile。用看树程序 TreeView 打开 treefile 文件，如图，可见不同的蛋白激酶被准确地归为不同的家族，而且有意思的是原癌基因 *ras* 等距离酪氨酸蛋白激酶 *src* 等的距离最近，提示它们在进化上的某种联系。



第五节　刊有生物信息学的期刊

随着人类基因组计划日新月异的发展，生物学与其他学科如医学、计算机科学等的交叉联系越来越广泛，在不久的将来，很难想像一个不懂计算机的分子生物学家能搞好科研（更不用说其水平如何）。所以对于从事这一相关领域的科研人员来说，了解生物信息学是时代发展的需要。为便于分子生物学家（而非计算机专家）提高生物信息学分析水平，特列出下列刊有生物信息学论文的较高水平的期刊。

1. *Nature*　生物信息学方面的文章较少，但不发则已，一发惊人。但其方法学部分一般过于简单，对初学者难度较大。尽管 Nature 水平可能是最高的杂志，但可能也是最小气的杂志之一，想免费看全文（哪怕是过刊），No way！所以没有必要列出其网址。

2. *Science*　相比之下，同样是最高水平的期刊，Science 显得比较大方，生物信息学方面的内容比较多，是科研工作者的必去场所（http://www. sciencemag. org/）。

3. *Cell*　生物信息学方面的文章较少，也属最小气的杂志之一，拟专研生物信息学的读者，在初期，不去也罢。

4. *Proc Natl Acad Sci Usa*　水平很高，可免费查全文。尤其是其 Developmental Biology、Evolution、Genetics 等栏目常有运用生物信息学分析的高水平文章。

5. *Trend Biochem Sci*　尽管是生化方面的综述类期刊，但几乎每期都有生物信息学分析的文章，生物信息学新方法，用计算机分析发现新的蛋白结构域等，但不免费。

6. *Genome Res*　水平较高，刊登比较基因组学、蛋白质组学、疾病基因克隆、进化分析和技术革新等许多方面的文章。过刊免费查阅（http://www.genome.org/）。

7. *Nucleic Acids Res*　水平不错，每年一月第一期为数据库专刊。过刊免费查阅（http://nar.oupjournals.org/）。

8. *Bioinformatics*　它是生物信息学的专门期刊，过刊免费查阅（http://bioinformatics.oupjournals.org/）。

（赵　勇）

第八章

酵母双杂交技术

内容提要

蛋白质与蛋白质的相互作用是细胞生命活动的基础和特征。酵母双杂交系统是一种研究蛋白质相互作用的强有力方法。该系统利用酵母转录激活因子 GAL4 的功能特点，通过两个杂交蛋白在酵母细胞中的相互作用及对报告基因的转录激活来捕获相互作用的蛋白质。本章介绍了酵母双杂交技术的原理与基本流程；实验中常使用的菌株、表达载体及培养基；与双杂交有关的预实验；在确证诱饵蛋白无毒性、无渗漏、无自激活的基础上，正式筛选预转化 cDNA 文库的具体操作步骤及注意事项；以及如何逐步筛选阳性克隆，并对其验证、归类和测序，从而最终得到与已知蛋白质相互作用的蛋白质。

随着人类后基因组时代的到来，蛋白质组学（proteomics）日益受到重视。在生命活动中，DNA 的复制、转录、RNA 的剪接、蛋白质的翻译和分泌以及细胞周期的调节、信号传导和诸多代谢过程都是各种相关蛋白质有机相互作用的结果。细胞或组织中的蛋白质不是杂乱无章的混合物，蛋白质间的相互作用、相互协调是细胞进行一切代谢活动的基础。蛋白质组学的研究目标是对机体所有的蛋白质进行结构鉴定和功能分析。以往研究蛋白质的方法多采用生物化学技术，如偶联、免疫共沉淀等方法，随着分子生物学实验技术的飞速发展，近 10 年来，发展了一些研究蛋白 - 蛋白相互作用的方法，其中以 Fields 和 Song 创立的酵母双杂交技术为最佳。该方法利用酵母生长速度快，且容易操作的特点，在其体系中研究哺乳动物细胞的蛋白 - 蛋白间的相互作用，通过筛选 cDNA 文库，找到与感兴趣蛋白相互作用的蛋白。本章就酵母双杂交原理、具体操作方法及一些值得注意的问题予以介绍。

第一节　酵母双杂交原理

一、原　　理

酵母双杂交系统的建立得力于对真核细胞调控转录起始过程的认识。研究发现，许多真核生物的转录激活因子都是由两个可以分开的、功能上相互独立的结构域（domain）组成的。例如，酵母的转录激活因子 GAL4，在 N 端有一个由 147 个氨基酸组成的 DNA 结合域（DNA binding domain，BD），C 端有一个由 113 个氨基酸组成的转录激活域（transcription activation do-

main，AD）。GAL4 分子的 DNA 结合域可以和上游激活序列（upstream activating sequence，UAS）结合，而转录激活域则能激活 UAS 下游的基因进行转录。但是，单独的 DNA 结合域不能激活基因转录，单独的转录激活域也不能激活 UAS 的下游基因，它们之间只有通过某种方式结合在一起才具有完整的转录激活因子的功能。

酵母双杂交系统正是利用了 GAL4 的功能特点，通过两个杂交蛋白在酵母细胞中的相互结合及对报告基因的转录激活来捕获新的蛋白质，其大致步骤为：

1. 视已知蛋白的 cDNA 序列为诱饵（bait），将其与 DNA 结合域融合，构建成诱饵质粒。

2. 将待筛选蛋白的 cDNA 序列与转录激活域融合，构建成文库质粒。

3. 将这两个质粒共转化于酵母细胞中。

4. 酵母细胞中，已分离的 DNA 结合域和转录激活域不会相互作用，但诱饵蛋白若能与待筛选的未知蛋白特异性地相互作用，则可激活报告基因的转录；反之，则不能。利用 4 种报告基因的表达，便可捕捉到新的蛋白质。

二、酵母双杂交的基本流程

见图 8－1。

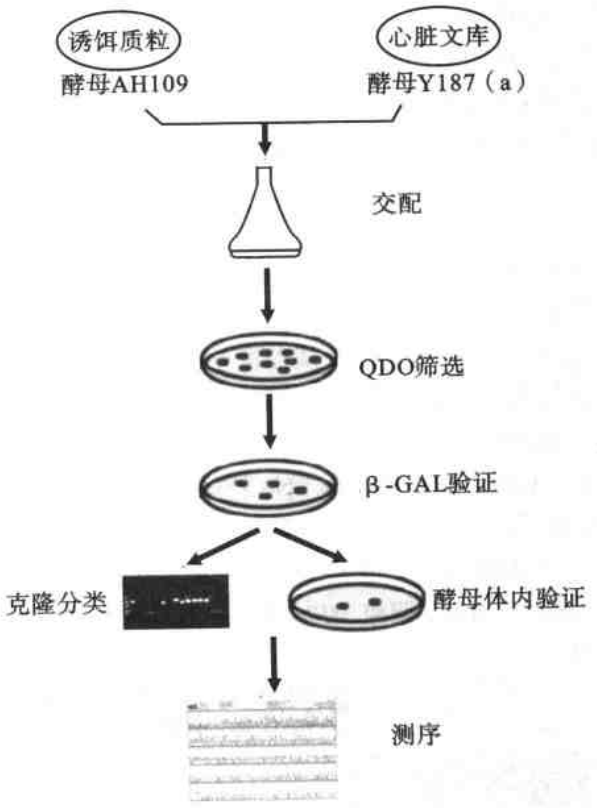


图 8－1　酵母双杂交的基本流程

注：首先构建诱饵质粒，即将已知的基因或基因片段克隆至载体中，并以此为诱饵去钓取

能与其相互作用的蛋白质，故称该质粒为诱饵质粒。再构建携带有大量信息的、待筛选的 cDNA 文库质粒，亦称文库质粒。分别将诱饵质粒转化入 AH109 酵母菌株、文库质粒转化入 Y187 酵母菌株后，将这 2 种带有不同基因的酵母菌置于同一培养瓶中培养(即交配培养，使它们形成二倍体细胞)，再将交配培养物铺于缺少某些营养成分的培养盘中。交配期间，如果诱饵质粒中已知基因所表达的蛋白质可以和文库质粒所表达的某些蛋白质发生相互作用，就会激活报告基因。利用这一特征，经 β－半乳糖苷酶(β－galactosidase)筛选，便可得到若干个可以和已知基因发生相互作用的阳性克隆。借助 PCR 扩增及酶切分析，将克隆进行分类。最后对阳性克隆进行测序，便可得到和已知蛋白发生相互作用的蛋白质。

第二节　酵母双杂交系统中常用的菌株与表达载体

一、酵母菌的一般特点

酵母菌是一种典型的真核生物，虽然其基因组较小，但其生物学特性与真核细胞相似。酵母菌是在 30 ℃条件下生长，在有氧和无氧状态中都可以很好地适应生存，它可以以出芽的方式进行有丝分裂，无限期生长。酵母菌还能以稳定的单倍体和二倍体的形式存在。单倍体细胞可以是两个交配型中的任意一个，二倍体细胞则是两个交配型的单倍体融合而成。在碳源和氮源不足的情况下，二倍体细胞只能以减数分裂的形式生存。酵母繁殖迅速，可以方便、廉价、大量地培养且无毒性。近年来随着对酵母体系的研究、改造、开发，不断有新的菌株产生，即使在酵母双杂交实验体系中，也在更新着更为合理、方便、稳定的菌株。为方便读者理解，本章根据 CLONTECH 公司 MATCHMAKER Libraries 2000 年产品中使用的菌株进行叙述。

二、常用的菌株

酵母双杂交系统中使用的菌株均已经过人工改造。它们含有特定的报告基因，并已删除了相应的转录因子编码基因，因此本身无报告基因的转录活性。当诱饵蛋白和待筛选的蛋白相互作用时，导致 BD 与 AD 在空间上接近，从而激活 UAS 的下游启动子，使报告基因 His 和 Leu 得到转录表达，因此转化子可在缺乏相应成分的培养基上生长，并由于 LacZ 的表达使其在含有 X－gal 的培养基上呈蓝色。

1. AH109 酵母菌株。AH109 酵母菌是诱饵质粒的宿主菌。它含有 4 个报告基因：*HIS*3、*ADE*2、*MEL*1 和 *LacZ*(也可将 *MEL*1 和 *LacZ* 归视为 1 个半乳糖苷酶报告基因)，它们都在 *GAL*4 UAS 的调控之下。最强的启动子是 *GAL*1，其次是 *GAL*2，*MEL*1 最弱。*GAL*1 和 *GAL*2 分别调控 *HIS*3 和 *ADE*2 的表达，*MEL*1 调控 *MEL*1 和 *LacZ* 的表达，二者又分别编码 α－半乳糖苷酶和 β－半乳糖苷酶(图 8－2)。

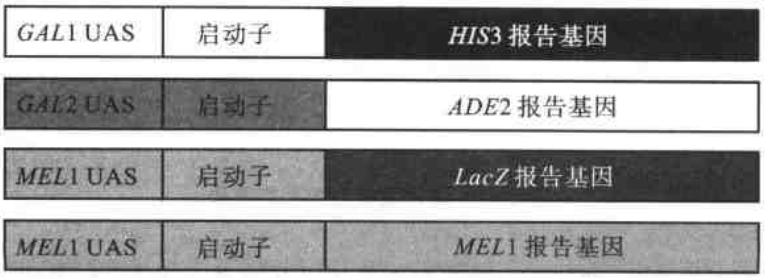


图 8－2　酵母菌株 AH109 的报告基因结构图

2. Y187 酵母菌株。Y187 酵母菌是文库质粒的宿主菌。它含有 2 个报告基因：*MEL*1 和 *LacZ*，分别由 *MEL*1 和 *GAL*12 个启动子调控（图 8－3）。

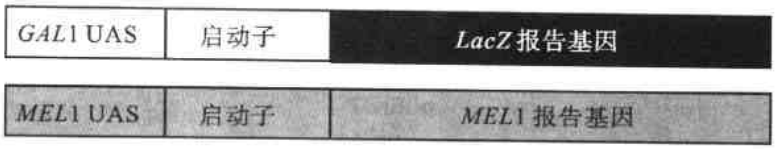


图 8－3　酵母菌株 Y187 的报告基因结构图

三、表达载体的构建

（一）诱饵质粒构建

将已知蛋白质的 cDNA 序列构建在某一含有 DNA 结合域序列的载体中，以其为诱饵，在庞大的 cDNA 文库中钓取能与之相互作用的蛋白，该质粒称之为诱饵质粒，为在实验中记录与叙述方便，写为：DNA－BD/bait。其特点是：它是一个融合表达载体，含有酵母转录因子 GAL4 的 DNA 结合域序列和两种可供选择之用的基因，一为抗生素抗性标记，另一为选择性营养标记，且均为显性表型。与常用于细菌质粒中的抗生素抗性标记所不同的是，选择性营养标记可以补偿酵母菌细胞中某一种特定的代谢缺陷（营养缺陷），即它必须与具有可被互补的经过适当突变的宿主菌株结合使用。由于该质粒可在大肠杆菌和酵母菌二种细胞中稳定存在并自主复制，故又称为穿梭质粒。

如质粒 pGBKT7（图 8－4），它含有卡那霉素抗性基因和可表达色氨酸（Trp）的选择性营养标记基因。卡那霉素抗性基因可用于细菌筛选，将质粒转化入细菌后，可在加有卡那霉素的培养基中生长。选择性营养标记基因用于酵母筛选，当质粒转化入酵母菌后，可在缺乏色氨酸的培养基中生长。诱饵质粒构建中要求将外源基因正确地克隆入表达载体，其起始密码正好位于启动序列的相应部位，以保持读码框一致。

（二）AD 融合的 cDNA 文库质粒的构建

为了能够寻找到与已知蛋白相互作用的蛋白，将待筛选的某种细胞、组织、器官或基因片段或基因突变体等构建成 cDNA 文库，并克隆入能够编码 AD 转录激活域的载体中，称为 AD 载体或 cDNA 文库载体，常用的有 pACT2（图 8－5）、pTD－1 等。该类载体同样为穿梭质粒，同样编码抗生素抗性标记和选择性营养标记，但均不同于诱饵质粒，如 pACT2 载体含有氨苄青霉素抗性基因和可表达亮氨酸（Leu）的选择性营养标记基因。随着分子生物学技术的商品化，酵母双杂交所用的各种 cDNA 文库可直接从公司购得，且公司已将构建的文库质粒转化在相应的酵母菌株中，故其商品名为预转化 cDNA 文库。如 CLONTECH 公司的人心脏 cDNA 文库已经预转化在 Y187 酵母菌株中，可直接用于筛选。

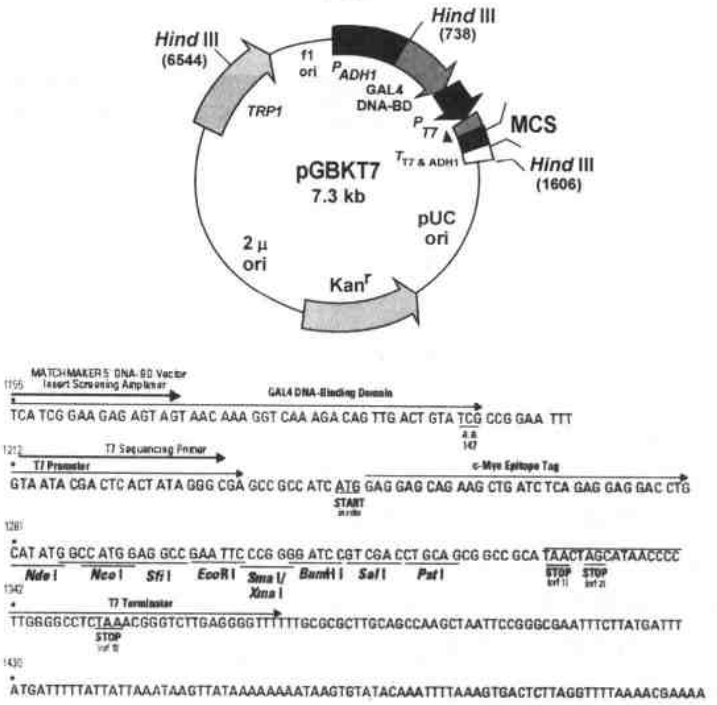


图 8－4　BD 融合的诱饵质粒 pGBKT7 的结构图及多克隆位点

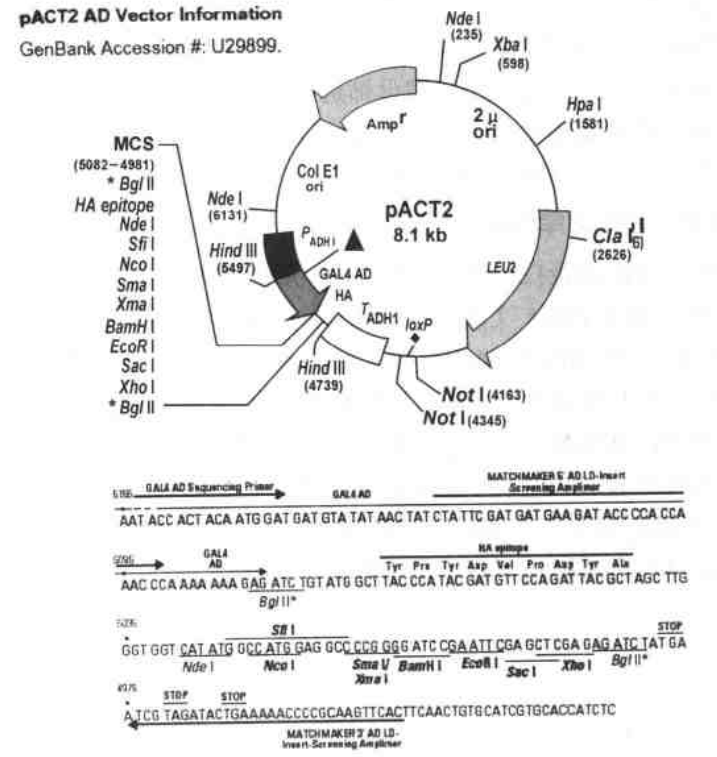


图 8－5　AD 融合的文库质粒 pACT2 的结构图及多克隆位点

四、表达载体与菌株表型之间的关系

酵母双杂交系统对相互作用蛋白的筛选过程，最终是通过利用诱饵 BD 质粒和文库 AD 质粒的选择性营养标记以及改造后的酵母菌株的报告基因的表达来实现的。简言之，含有表达 Trp 蛋白基因的诱饵 BD 质粒可在缺乏 Trp(SD/－Trp)的培养基中生长；含有表达 Leu 蛋白基因的文库 AD 质粒可在缺乏 Leu(SD/－Leu)的培养基中生长。当二者共同转化至酵母形成二倍体细胞后，就可在缺乏 Trp 和 Leu(SD/－Trp/－Leu)的培养基中生长；当诱饵蛋白和待筛选蛋白发生相互作用后，BD 与 AD 接近，激活了 UAS，使各报告基因得以转录，因此可在 SD/－Ade/－His/－Leu/－Trp 的培养基中生长，并由于 LacZ 或 MEL1 基因的表达，通过 x－gal 显色筛选即可得到阳性克隆。

五、对照载体的构建

酵母双杂交中，设计好阳性对照和阴性对照是十分重要的。在 CLONTECH 公司所有的 MATCHMAKER Libraries Kit 中都提供了一对现成的阳性对照 BD 质粒 pGBKT7/53(图 8－6)和 AD 质粒 pTD1－1(图 8－7)，它们分别携带有鼠源 *P53* 基因和 SV40 大片段 T 抗原，并已分别转化入 AH109 和 Y187 酵母菌株中。已经肯定两个基因可以发生相互作用，当二者交配为二倍体细胞后，在 SD/－Leu/－Trp 培养基中可以生长，若经 β－gal 检测克隆均应呈蓝色，且不会有渗漏与自激活。当然，该二倍体细胞也可在 TDO 和 QDO 培养基中生长，但一般不必检测。

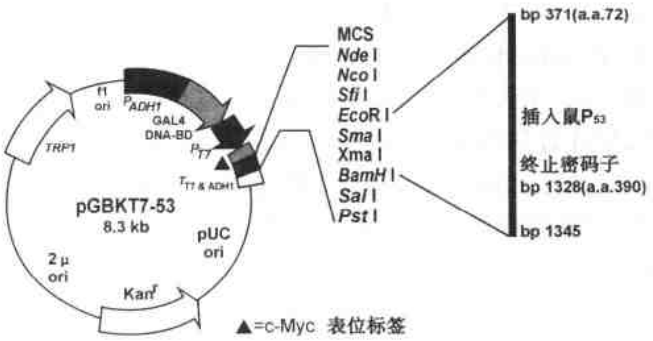


图 8－6　阳性对照 BD 质粒 pGBKT7－53 的结构图

阳性对照 AD 质粒 pTD1－1，是一特异设计的 AD 载体，除与阳性对照 BD 质粒 pGBKT7/53 外，不与其他诱饵质粒发生相互作用。即使与其他诱饵质粒交配成二倍体细胞，也不会激活 GAL4 上游激活系统的，因此，无论在 TDO 或 QDO 培养基中都不会生长。反之，若在 TDO 或 QDO 培养基中有克隆生长，则证明诱饵质粒有自激活。AD 质粒 pTD1－1 除用于阳性对照和阴性对照外，还用于检测诱饵蛋白对酵母交配效率的影响。

阴性对照 BD 质粒往往以空载体如 pGBKT7 或 pGBKT7/Lam 为诱饵质粒，前者不含有任何外源基因片段，后者虽可编码一个融合的人核纤层蛋白 C，但它基本不与作为阳性对照的 AD 质粒 pTD1－1 或其他的待筛选文库质粒发生相互作用。将 pGBKT7 或 pGBKT7/Lam 转化入 AH109 酵母后均可与 Y187/pTD1－1 交配成二倍体细胞，仅能在 SD/－Leu/－Trp 培养基中生

长克隆。

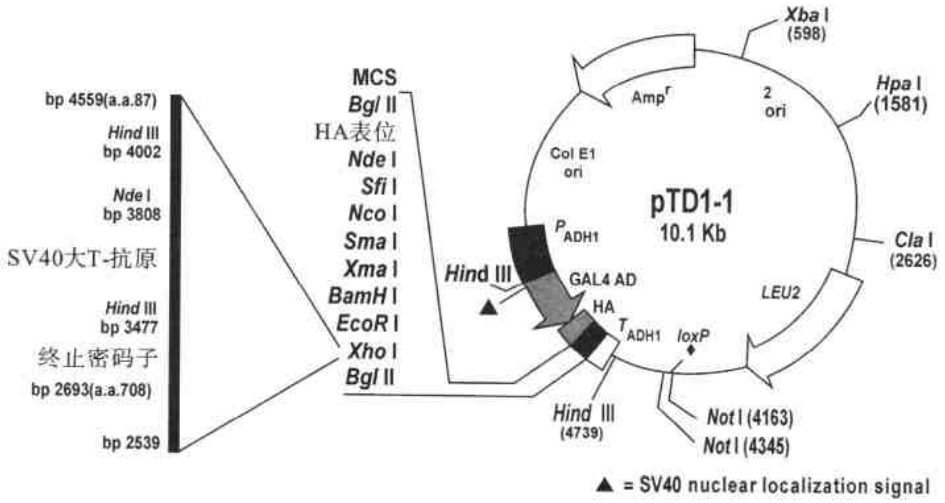


图 8－7　阳性对照 AD 质粒 pTD1－1 结构图

六、酵母培养基的制备

酵母菌在液体培养基和固体培养基中均可生长，同时还可在营养丰富和营养缺陷的培养基中生长。营养丰富培养基，是培养酵母菌最常用的一般性培养基，主要组分是酵母提取物(Yeast extract)、蛋白胨(Peptone)和葡萄糖(Dextrose)，由于这种培养基提供了细胞从头合成所需的氨基酸、核苷酸前体、维生素和其他代谢物，常称为丰富培养基或 YPD 培养基。营养缺陷的培养基，是用于培养已转化过的酵母菌的最基本培养基，以选择和检测它们的表型，常用于培养交配型酵母菌或用于选择二倍体、选择转化子酵母菌等。它包括极限培养基(亦称 SD 培养基)和完全极限培养基(亦称选择性压力培养基、缺陷型培养基)。

(一)液体培养基

1.丰富培养基。

(1) YPD 培养基：

蛋白胨(Peptone) 20 g

酵母提取物(Yeast extract) 10 g

加 H_2O 至 950 mL，高压灭菌。待培养基冷至 55 ℃时加 40%经过滤的葡萄糖至终浓度为 2%。

(2) YPDA 培养基：在 1 LYPD 培养基中加入 15 mL 过滤灭菌的 0.2% 腺嘌呤半硫酸盐(终浓度为 0.003%)。

2.极限培养基(亦称 SD 培养基)。是培养酵母菌的最基本培养基，它是由以葡萄糖为碳源、以无机盐提供氮源、磷元素及微量金属和 10×省却成分培养基所组成。

配制：无氨基酸的酵母基本氮源(yeast nitrogen base without amino acids，YNB，Sigma 公司产品)6.7 g，加水至 850 mL，高压灭菌。待培养基冷至 55 ℃时加 40%经过滤的葡萄糖至终浓度为 2%，100 mL 灭菌的含有 13 种成分的 10× DO 培养基。根据需要，还可加 3－AT、放线菌酮

(cycloheximide)、腺嘌呤半硫酸盐或 X－gal 等。

3. 10×省却成分培养基(dropout supplements，简称 10×DO 培养基)。

L－异亮氨酸（L－Isoleucine）	300 mg/L
L－缬氨酸（L－Valine）	1 500 mg/L
L－腺嘌呤半硫酸盐（L－Adenine hemisulfate salt，Ade）	200 mg/L
L－精氨酸（L－Arginine HCl）	200 mg/L
L－组氨酸（L－Histidine HCl monohydrate，His）	200 mg/L
L－亮氨酸（L－Leucine，Leu）	1 000 mg/L
L－赖氨酸（L－Lysine HCl）	300 mg/L
L－蛋氨酸（L－Methionine）	200 mg/L
L－苯丙氨酸（L－Phenylalanine）	500 mg/L
L－苏氨酸（L－Threonine）	2 000 mg/L
L－色氨酸（L－Tryptophan，Trp）	200 mg/L
L－酪氨酸（L－Tyrosine）	300 mg/L
尿嘧啶（L－Uracil）	200 mg/L

高压灭菌，4 ℃保存，或随 SD 溶液一并消毒。

10×省却成分培养基共含有 13 种成分，且 10 倍浓度。习惯上，若只写 10×DO，是指含有 13 种成分的溶液，并不缺少任何成分。真正的省却之意，是指该液体中缺少了某一种或几种成分，但含有其他成分，并仍为 10×浓度。常用的省却了某种成分的 10×DO 包括以下几种形式。

一缺 10×DO，有 2 种：10×DO/－Trp、10×DO/－Leu

二缺 10×DO，有 3 种：10×DO/－Leu/－Trp、10×DO/－Ade/－Trp、10×DO/－His/－Trp

三缺 10×DO，有 2 种：10×DO/－Ade/－Leu/－Trp、10×DO/－His/－Leu/－Trp

四缺 10×DO，只有 1 种：10×DO/－Ade/－His/－Leu/－Trp

它们分别代表缺少的成分。各种 10×DO 最好事先单独配制且标记清楚，以防混乱。省却成分培养基的用途在于检测和筛选。

4. 完全极限培养基(亦称缺陷型培养基、选择性压力培养基)。配制缺陷型培养基时，需在 SD 溶液的基本配伍(YNB 6.7 g，加水至 850 mL，40%葡萄糖加至终浓度为 2%）中加入 1/10 体积的、不同省却成分的 10×DO 培养基。按其省却的种类，缺陷型培养基大约分为 4 类，并根据省却的成分给予相应的冠名：缺少一种成分的培养基，有 SD/－Trp 和 SD/－Leu；缺少二种成分的培养基，有 SD/－Leu/－Trp、SD/－Ade/－Trp 和 SD/－His/－Trp，统称 DDO 培养基；缺少三种成分的培养基，有 SD/－His/－Leu/－Trp 和 SD/－Ade/－Leu/－Trp，统称 TDO 培养基；缺少四种成分的培养基，有 SD/－His/－Ade /－Leu/－Trp，称作 QDO 培养基。

注意：①SD 培养基是指加入含有 13 种成分的 10×DO；选择性压力培养基是缺陷型培养基，即加入缺少了不同成分的 10×DO；②配制培养基的各种试剂尽可能使用相对固定的公司产品，尤其是对 10×省却成分培养基中的 13 种试剂要从同一公司购买(推荐使用 Sigma 公司产品)，且数量配置要合适；③一般情况下 SD 缺陷和非缺陷型培养基不需加抗生素；④实验室内应统一对培养基的名称，标签与记录一致。

(二) 固体培养基

无论对哪一种液体培养基，加入 2%(20 g/L)的琼脂粉，高压灭菌，冷却后铺盘就成为相应

的固体培养基了。铺盘前，待培养基冷却至 55 ℃时加入待添加的试剂和营养成分（如：葡萄糖、腺嘌呤、3－AT 或 X－α－Gal 等），充分混匀，尽量避免产生气泡，在超净台内倒入培养皿。每升培养基可以制备直径 150mm 的培养盘 13～14 个（约 70 mL/个，过少易干裂）；制备直径 90 mm 的培养盘 33～38 个（约 30 mL/个）。应在使用前制备，4 ℃下可保存 2 个月。

说明：本章中所写培养皿一词是指未加入任何培养基，加入固体培养基后的培养皿统称培养盘，以示区别。

（三）相关试剂的制备

1. 40%葡萄糖：葡萄糖 40 g，加水至 100 mL，过滤后 4 ℃保存（因液体粘稠，加压过滤时应小心滤膜破损）。

2. 50 mg/mL 的卡那霉素（kanamysin）：过滤后分装，－20 ℃保存。当在某种培养基（如 YPDA）中加入卡那霉素时，可简写作：YPDA/kan。

3. 1M 3－AT：1M 3－AT（3－amino－1,2,4－triazole，3－氨基－1,2,4－三偶氮），因名字冗长，中、英文文献中均简称 3－AT）溶液，过滤除菌，4 ℃保存。因 3－AT 易热灭活，故使用时须待培养基温度降至 55 ℃以下再加入。

4. X－α－Gal 储备液：X－α－Gal 即 X－α－半乳糖吡喃糖苷，全称：5－bromo－4－chloro－3－indolyl－α－D－galactopyranoside。它是 α－半乳糖苷酶的底物，在酵母双杂交系统中用于检测报告基因 MEL1 的表达。α－半乳糖苷酶是 MEL1 的表达产物，它是一分泌型酶，因此可直接在 X－α－Gal 培养盘上产生显色反应，克隆呈蓝色。X－α－Gal 可先溶解于培养基中再铺盘，也可涂于培养基表面，较 β－gal 滤纸分析法方便，但其价格较昂贵。

制备：将 X－α－Gal（CLONTECH 公司）溶于二甲基甲酰氨（dimethylformamide，DMF）中，浓度 20 mg/mL。－20 ℃，避光保存。使用时每 100 mL 培养基中加 100 μL X－α－Gal 储备液。

5. 各种贮存液：

（1）50% PEG 3350：称取 PEG 3350 50 g（很轻，但占的体积很大），用灭菌的去离子水溶解，至体积达 100 mL。如不易溶解，可加热至 50 ℃助溶。

（2）10×TE 缓冲液：0.1M Tris·HCl，10mM EDTA，pH7.5，高压灭菌。

（3）10×LiAc：1M LiAc，用稀释的醋酸调至 pH7.5，高压灭菌。

（4）100% DMSO。

6. PEG/LiAc 溶液（现用现配）：制备使用液 10 mL。取 8 mL 50%PEG、1 mL 10×TE 缓冲液和 1 mL 10×醋酸锂（LiAc）混匀，备用。

七、菌株冻存与复苏

1. 冻存液：指液体培养基中加入终浓度为 25%的灭菌甘油。

2. 菌株冻存原则：未经转化的酵母菌株多采用 YPD 培养基进行冻存，但 AH109 菌株需用 YPDA 培养基冻存；对转化了各种质粒的酵母菌进行冻存，为保持质粒的选择压力，应使用相应的 SD（或 YPD）培养液，即其中 10×DO 溶液的种类要依据实验而定。一般液体、固体培养基及冻存液应保持一致。冻存时，或挑取在培养盘中正值生长旺盛的单克隆置入事先放好冻存液的 Eppendorf 管中，震摇混匀以彻底分散细胞；或吸取在液体培养基中正值对数生长期的菌液，加入甘油混匀，置－70 ℃保存。冻存期可达 1 年以上。

3. 复苏：事先铺盘。未经转化的酵母菌株复苏时，使用 YPD 或 SD 盘，但 AH109 菌株要使用 YPDA 盘。对转化了各种质粒的酵母菌复苏时，可使用 YPD 或 YPDA，只要孵育时间较短，

质粒不会丢失即可。但若要从转化的酵母中表达融合蛋白，则应按照冻存时的状态使用相应省却成分的 SD 盘。

复苏时，在冻存管的表层（不要溶化整个冻存管），以灭菌牙签蘸取小量细胞，在盘上划线。倒置于 30 ℃温箱培养，3～5 d 后，长出克隆，直径约 2 mm。以封口膜封闭培养盘，可置 4 ℃存放 2 个月。

八、酵母菌的培养

酵母菌在 30 ℃，氧气充足和以葡萄糖为碳源并加入无机盐提供氮源、磷元素及微量金属的极限培养基（基本培养基）中即可生存，若在培养基中加入蛋白质和酵母细胞提取物（丰富培养基），酵母细胞可以生长得更快些。在双杂交实验中，如果欲培养单纯的未经转化的菌株，应使用 YPD 或 YPDA 营养丰富培养基，若为培养交配型酵母菌或用于选择二倍体、选择转化子酵母菌等，应使用 SD 和/或 SD 缺陷型培养基（视实验而定）。以液体培养基大量培养酵母菌时，应使用锥形瓶，且培养液不能超过烧瓶体积的五分之一。为保证良好的通气条件，瓶口不应完全封闭。培养瓶应置于 230～300 r/min，30 ℃恒温摇床中进行培养。使用固体培养基时，每个直径 150 mm 的培养皿中，培养基的量不应少于 70 mL；每个直径 90 mm 的培养皿中，培养基的量不应少于 25 mL。否则，酵母菌营养不良，克隆生长的体积小，且易干裂。

第三节　诱饵质粒的鉴定

诱饵质粒构建完成后，需进行一系列鉴定实验，常称之为预实验。其主要内容包括：诱饵质粒转化入感受态酵母细胞、诱饵蛋白毒性作用检测、诱饵蛋白渗漏与自激活作用检测及小规模酵母交配实验等。

一、制备感受态酵母菌

（一）材料准备

诱饵质粒、空载体 pGBKT7 或阴性对照载体 pGBKT7/Lam、阳性对照载体 pGBKT7/p53（若尚未转化入酵母）、AH109 菌株、YPDA 液体培养基、SD/－Trp 培养盘、灭菌的 1×TE 缓冲液、1×醋酸锂（LiAc）溶液、PEG/ LiAc 溶液、DMSO、10 mg/mL 鲑精 DNA、灭菌的直径 5 mm 玻璃珠。

（二）感受态酵母菌的制备

以 AH109 酵母菌为例。

1. 从复苏的 AH109/YPDA 盘上以牙签挑取 1 个直径在 2～3 mm 的克隆接种于装有 1 mL YPDA 液体培养基的 Eppendorf 管中。若克隆直径小于 2 mm，应挑取数个克隆。

2. 剧烈振荡 5 min 以打散菌团。

3. 将之转入含 50 mL YPDA 培养基的三角瓶中。

4. 于 30 ℃，250 r/min 摇动 10～18 h 至酵母菌繁殖的稳定期（stationary phase）（OD_{600}值＞1.5）。

5. 取 30 mL 过夜培养物转入含 300 mL YPDA 的三角瓶中，检查稀释培养物的 OD_{600}值，若必要，加入更多的过夜培养物直至 OD_{600}至 0.2～0.3。

6. 于 30 ℃，230 r/min 继续摇动 3 h，此时 OD_{600}应为 0.4～0.6。

7. 收集培养菌液于 50 mL 离心管中，1 000 g，室温（20～24 ℃）离心 5 min。

1 2 3 5 6 7 8 甲乙丙甲乙丙

8. 弃上清，加 25～50 mL 灭菌 TE 或蒸馏水，震摇，彻底悬起菌体团块。

9. 1 000 g，室温离心 5 min。

10. 弃上清。

11. 将菌体团块悬浮于 1.5 mL 新鲜准备的、灭菌的 1×TE/1×LiAc（各为 50% 体积）溶液中。

注意：应尽可能在超净台中操作；随时检测酵母的 OD 值，防止超过对数生长期。

二、诱饵质粒转化入感受态酵母细胞

（接上面步骤）

12. 取 0.1 μg 待转化的诱饵质粒 DNA 和 0.1 mg 鲑精 DNA 于一新的 1.5 mL 离心管中，混匀。

13. 加 0.1 mL 酵母感受态细胞，震摇混匀。

14. 加 0.6 mL 灭菌 PEG/LiAc 溶液，并高速震摇 10 s 以混匀。

15. 30 ℃，200 r/min 摇动 30 min。

16. 加 70 μL DMSO，轻轻颠倒混匀，切勿震摇。

17. 42 ℃水浴中热激 15 min。

18. 冰浴 1～2 min。

19. 14 000 r/min，室温离心 5 s，移去上清。

20. 用 0.5 mL 灭菌 1×TE 缓冲液悬浮细胞（为叙述方便，将已转化了质粒的酵母菌称为转化子）。

21. 取 100 μL 细胞悬浮液滴加在 SD/－Trp 培养盘表面，放入 4～6 颗灭菌玻璃珠，水平晃动以使细胞悬液均匀铺展，这对克隆的均匀生长非常关键。平放几分钟，至液体渗入琼脂中为止。

22. 为保证获得分离良好的克隆，同时为作为计算共转化效率的对照，再以 1×TE 缓冲液制备 1∶10、1∶100 和 1∶1 000 的稀释细胞悬浮液，重复第 21 步操作。

23. 于 30 ℃培养箱倒置培养，直至克隆出现。一般需要 2～4 d。

24. 酵母克隆明显大于细菌克隆，直径至少应达 2 mm。将培养盘用封口膜封好，4 ℃可存放 3～4 周。

25. 为了得到更好的分离，以确证转化子的正确，还可以挑取最大克隆，于同样的选择培养基中再次划线及培养以作为储备板保存。

注意：此处仅以将诱饵质粒转化入 AH109 为例，若转化其他质粒或为其他目的的转化，则应准备相应的 SD 缺陷型培养盘。

三、诱饵蛋白的毒性检测

（一）原理

通过比较转化了空质粒 pGBKT7 和转化了诱饵质粒 DNA－BD/bait 的酵母菌在液体培养基中的生长速率，以了解 DNA－BD/bait 蛋白是否有毒性作用。正常状态下，二者的生长速率相同或接近，如果后者的生长速率明显较前者慢，说明 DNA－BD/bait 蛋白可能对酵母菌有毒性作用，因此需将酵母铺在琼脂盘上与预转化文库交配，通过交配效率的计算来决定该诱饵质粒是否可继续使用。

(二)方法

1. 分别挑取转化了空质粒 pGBKT7 和诱饵质粒的大个单克隆，放入装有 5 mL SD/－Trp 液体培养基的摇菌管中，振荡混匀，将摇菌管以相同倾斜度放置，于 30 ℃，200 r/min 振摇培养 15 h。

2. 次日取 2 mL 过夜菌移至 20 mL 相同的培养基中，30 ℃，250 r/min 继续振摇培养 3 h。测 OD_{600}值，进行比较。

四、诱饵蛋白自激活检测

(一) 常用名词

转化有质粒的酵母菌，不论质粒的性质如何，常称其为转化子、单倍体细胞；一旦两种单倍体细胞交配，就称为二倍体细胞。在酵母双杂交体系中，单克隆的概念主要是指由克隆本身生长形成的、形体上与周围克隆互不相连的、独立的酵母单菌落。但就其本质而言，它可以是单倍体细胞，也可以是二倍体细胞。

(二) 渗漏与自激活现象

转化有 DNA－BD/bait 质粒的 AH109，在与 DNA－AD 文库质粒交配前，是不能转录激活报告基因 HIS3 的表达的，即只能在 SD/－Trp 培养基上生长，而不应在 SD/－His/－Trp 培养基上长出克隆，一旦有克隆生长，就称为渗漏现象。其产生的原因可能是构建在诱饵载体中的外源基因片段本身就是一个转录因子或者具有转录因子的作用，从而激活了报告基因。3－AT 是 HIS3 蛋白的竞争性抑制剂，它用来抑制 HIS3 蛋白的表达。如果渗漏现象不严重，且能够被 3～15 mM 浓度的 3－AT 抑制，那么该诱饵质粒可继续用于筛选。

同理，也不应有其他报告基因的表达。如以 *MEL*1 为代表检测有无半乳糖苷酶报告基因的表达：制备加有 X－α－gal 的缺陷型 SD 培养盘，不论缺陷的成分与种类，只要有蓝色克隆生长，即为激活了 *MEL*1 基因，此时称为自激活。因此，在正式筛选之前，必须进行 X－α－gal 检测，以确证诱饵质粒不会激活报告基因，以防止或降低假阳性克隆的产生。

渗漏与自激活是不同的。前者(若不严重)可被 3－AT 所抑制，而后者则无抑制方法。

(三) 预实验需用下列酵母克隆

AH109[pGBKT7－＊＊＊]：代表研究者所构建的诱饵质粒，生长于 SD/－Trp 培养盘。

AH109[pGBKT7－53]：生长于 SD/－Trp 培养盘。

AH109[pGBKT7]：生长于 SD/－Trp 培养盘。

Y187[pTD1－1]：生长于 SD/－Leu 培养盘。

AH109[pGBKT7－＊＊＊]×Y187[pTD1－1]二倍体：生长于 SD/－Trp/－Leu 培养盘。

AH109[pGBKT7－53]×Y187[pTD1－1]二倍体：生长于 SD/－Ade/－His/－Trp/－Leu 培养盘(阳性对照)。

AH109[pGBKT7]×Y187[pTD1－1]二倍体：生长于 SD/－Trp/－Leu 培养盘(阴性对照)。

注："＊＊＊"代表已知蛋白的 cDNA 序列。当两个单倍体间用"×"连接时，表示二者已交配为二倍体细胞。

(四) 单倍体细胞的自激活检测

1. 若仅检测诱饵质粒，使用直径 90 mm 培养皿，制备 SD/－Trp/X－α－Gal，SD/－His/－

Trp /X－α－Gal 和 SD/－Ade/－Trp/X－α－Gal3 种选择性压力培养盘，视检测的质粒种类（每种用一个盘），决定铺盘量数。

2. 为防止渗漏现象出现和节省时间，可同时制备终浓度分别为 3、6、9、12、15mM 3－AT 的 SD/－His/－Trp/X－α－Gal 盘（3－AT 的梯度浓度，可自行调整）。

3. 在检测诱饵蛋白毒性后，用剩余的菌液作 1∶5 000 稀释，各取 200 μL 铺在上述培养盘中。

4. 晾干后，30 ℃倒置培养 3～5 d，待克隆长出。

5. 结果观察：诱饵质粒转化子应仅能在 SD/－Trp/X－α－Gal 盘上生长克隆，且为白色。若为蓝色，或在 SD/－Ade/－Trp/X－α－Gal 盘上有克隆生长，均为有相应报告基因的自激活，应分析原因。若在 SD/－His/－Trp/X－α－Gal 盘上出现白色克隆，为有渗漏。若有蓝色克隆出现，说明不仅有渗漏（即激活了 *HIS*3 报告基因），且还有 *MEL*1 基因的自激活。对有白色克隆生长的菌株，可用不同浓度的 3－AT 予以抑制。当从某一浓度的 3－AT 培养基开始无克隆生长（亦称无背景生长）时，为 3－AT 使用的最低浓度，该浓度可用于正式筛选。

6. 若转化的克隆为白色，还应制备存储培养盘（即在相同的缺陷型培养盘上划线生长，但不必加 X－α－Gal）或挑取白色克隆以冻存。

（五）二倍体细胞的自激活检测

1. 二倍体细胞的交配产生：二倍体可以由两个单倍体在液体培养基中交配产生，也可在固体培养基中用划线交配法产生。如：用牙签分别挑取 SD/－Trp 培养盘上的 AH109[pGBKT7－53] 克隆和在 SD/－Leu 培养盘上生长的 Y187[pTD1－1] 克隆，共同接种在 SD（非缺陷型培养基）培养盘上的同一位置并稍加混匀，即两种克隆重叠接种。30 ℃，交配 4 h 后，用牙签挑取交配物在 SD/－Ade/－His/－Trp/－Leu 培养盘上划线。30 ℃倒置培养 4～6 d。

2. 分别挑取实验组二倍体、阳性对照二倍体及阴性对照二倍体的单一克隆至有 1 mL 灭菌水的 Eppendorf 管中，振荡混匀，以 1∶5 000 稀释后取 100 μL 涂布下列培养盘。若有 HIS 表达渗漏，则需加 3～15mM 3－AT 以抑制渗漏。为检测 *MEL*1 报告基因表达，也可在铺交配混合物前将 X－α－Gal 加到选择培养基中。铺盘：

二倍体：SD/－Trp/－Leu/X－α－Gal、SD/－His/－Trp/－Leu、SD/－Ade/－Trp/－Leu，SD/－Ade/－His/－Trp/－Leu

阳性对照：SD/－Ade/－His/－Trp/－Leu/X－α－Gal

阴性对照：SD/－Trp/－Leu/X－α－Gal、SD/－His/－Trp/－Leu，SD/－Ade/－Trp/－Leu、SD/－Ade/－His/－Trp/－Leu

3. 结果观察：由 AH109[pGBKT7－＊＊＊] 和 Y187[pTD1－1] 交配的二倍体只能在 SD/－Trp/－Leu/ X－α－Gal 中生长，且为白色克隆，若在其他培养基中生长或克隆呈蓝色，说明诱饵质粒有渗漏和自激活；阳性对照二倍体在各种涂有 X－α－Gal 的选择性压力培养基 DDO、TDO 及 QDO 中均可生长，且为蓝色克隆；阴性对照也只能在 SD/－Trp/－Leu/X－α－Gal 中生长，并为白色克隆。

五、交配效率的检测与计算

为了能在正式筛选前对交配效率有一梗概的了解，尤其是了解研究者所构建的诱饵蛋白对酵母交配效率的影响，应进行小规模酵母预杂交实验，以检测和计算交配效率。

1. 分别从正在实验的储存培养盘中挑取体积大于 2～3 mm，且应新鲜（少于 2 周）的、大

小相近的一对单倍体克隆，如：AH109［pGBKT7－＊＊＊］和 Y187［pTD1－1］，同时放入一含 0.5 mL 2×YPDA 培养液的 EP 管中。

2. 振荡 1 min 以彻底悬浮细胞。

3. 随后将其转至 2 mL 2×YPDA 的摇菌管中。

4. 30 ℃ 200 r/min 振摇过夜（20～24 h），此期间为交配过程，产生二倍体细胞。

5. 次日铺盘：用 0.5×YPDA 以 1∶200 稀释交配混合物，并取 200 μL 涂布于下列两种 SD 选择培养基：SD/－Trp 和 SD/－Trp/－Leu。

注意：此处用 0.5×YPDA 培养基的目的在于少加些营养物，以提高二倍体的存活率而不会显著增加选择培养基的背景。

6. 30 ℃，倒置培养 4 d。

7. 计数两种培养盘上的克隆数。SD/－Trp 盘上的克隆数肯定多于 SD/－Trp/－Leu，因为只有二倍体细胞才能够在 SD/－Trp/－Leu 盘上生长。克隆数在 30～300 之间易于计数，因此要求铺盘时一定要有不同的稀释度。

8. 交配效率的计算：公式：二倍体细胞克隆数（SD/－Trp/－Leu 盘的克隆数）/单倍体细胞克隆数（SD/－Trp 盘的克隆数）×100 % = %交配效率

第四节　正式筛选预转化 cDNA 文库

一、筛选前的准备

（一）实验材料

筛选过程中需使用大量的培养皿，至少应准备 80 个直径 150 mm、200 个直径 90 mm 的培养皿、1 个 2 L 容积的三角瓶、若干个 0.5～1 L 容积的三角瓶、100 只摇菌管、1 kg 直径 5 mm 的玻璃珠、100 g 直径 0.4～0.5 mm 的玻璃珠以及 5 000 个左右的牙签，除后 2 种为一次性物品外，其余应可重复消毒使用。酵母菌是在 30 ℃条件下生长，筛选过程至少连续 2 周，故需单独的温箱，但不需特殊气体供应。

（二）文库的保存

文库应保存在－70 ℃。从公司购买的文库，多以 1 mL/支为一次使用量。文库液滴一旦融化，切勿再次冻存。

（三）制备培养盘

1. 筛选严谨程度的确定。

（1）严谨程度：在此是指在何种选择性压力培养基上进行筛选。

（2）高严谨选择：用缺乏 4 种营养物的培养基（即 SD/－Ade/－His/－Leu/－Trp，QDO）进行杂交的克隆筛选。可同时在培养盘上加 X－α－Gal 以检测 *MEL*1 报告基因的表达。

（3）低严谨选择：用缺乏 3 种营养物的培养基（即 SD/－His/－Leu/－Trp，TDO）进行杂交的克隆筛选。若需要可在培养盘中加 3－AT。

（4）折中严谨选择：将交配培养物的一半铺于 QDO 培养盘，另一半铺于 TDO 培养盘。

根据预实验的结果，决定正式筛选拟采用的严谨程度。当确信诱饵蛋白无明显毒性、无渗漏、无自激活，且预转化的交配效率在正常范围时，以使用高严谨选择性压力培养基为最佳方案。当然，此方法筛选所得到的阳性克隆数少于用低严谨和折中严谨法筛选的克隆，但假阳性

克隆相对也少。由于试剂 X－α－Gal 价格昂贵且需要量大，可用两次 X－β－Gal 检验来筛选和确证阳性克隆，以代替 X－α－Gal 的作用。培养基中高浓度的 3－AT 能杀死新鲜转化的细胞，所以当用过量 3－AT 来选择强的双杂合相互作用时，最好用低严谨选择方案。

2. 制备培养盘。无论拟采用何种压力选择性培养基，均需铺制 60～63 个直径 150 mm 的培养盘。同时制备直径 90mm 的 SD/－Leu，SD/－Trp，SD/－Leu/－Trp 培养盘各 3 个。严格无菌操作。虽然国外公司认为制备好的培养盘可在室温放置 2～4 个月，但据作者经验，以不超过 3 周为限。琼脂表面雾气水珠会导致细胞的不均匀分布，所以若培养盘存放在 4 ℃，当正式转化前应室温放置 2～3 d 或 30 ℃放置 3 h 后使用。

二、正式筛选预转化的 cDNA 文库

（一）筛选步骤

1. 准备浓缩的诱饵质粒宿主菌株的过夜培养物，如 AH109[pGBKT7－53]。

（1）接种一大（2～3 mm）的、新鲜（＜2 月）的 AH109[bait]克隆于 50 mL 的 SD/－Trp 培养基，30 ℃，250～270 r/min 振荡过夜（16～24 h）。

（2）次日检查培养物的 OD_{600} 值。在进行步骤 3 前，应 OD_{600} 值＞0.8。

（3）将 50 mL 培养物于 1 000 g 离心 5 min，弃上清，用剩余的 5 mL 液体中悬浮细胞团块。用计数板进行细胞计数，浓度应＞1×10^9 cells/mL。

2. 临用前于室温水浴融化约 1 mL 文库，轻轻振荡并取出 5 μL 置于冰中，以备随后文库滴定（先向下进行，待文库交配实验后再滴定文库）。

3. 在 2 L 容积的灭菌三角瓶中加入步骤 1 制备的 AH109[bait]培养物和 1 mL 文库，再加 45 mL 2×YPDA/Kan 并轻轻摇动。用 1 mL 2×YPDA/kan 冲洗冻存文库管 2 次。最后用 2×YPDA/kan 将终体积补至 50 mL。

4. 于 30 ℃ 30～50 r/min 轻轻振荡 20～24 h。

注意：三角瓶体积至少为 2 L，以使在低速振荡时交配混合物有足够的通气空间。低速振荡对保持细胞集中于瓶底，防止飞溅和贴壁是非常重要的，当振荡速度＞50 r/min 时将显著降低交配效率。

5. 在酵母交配 20 h 后，于一相差显微镜下（400×）检查一滴交配混合物。如见合子（zygotes），应继续交配 4 h。合子一般呈三叶形（three－lobed shape），各叶代表 2 个源自双亲的单倍体细胞和出芽的二倍体细胞（the budding diploid cell）（图 8－8）。

6. 终止交配。将交配混合物移至一灭菌的 100 mL 离心管，1 000 g 离心 10 min 沉淀细胞。随后用 2×YPDA/kan 冲洗交配用三角瓶 2 次，每次 50 mL。用两次冲洗液悬浮最初得到的细胞团块，再次 1 000 g 离心 10 min 沉淀细胞。最后，用 10 mL 0.5×YPDA/kan 悬浮细胞团块，并测量细胞和培养基的总体积（例：10 mL 培养基＋2.5 mL 细胞，共 12.5 mL。用于计算所筛选的克隆数）。

7. 交配效率测定：取 5 μL 文库交配混合物，用 0.5×YPDA/kan 做 1∶100，1∶1 000，1∶10 000 稀释，分别取各种稀释液 100 μL 涂布于 SD/－Leu、SD/－Leu/－Trp、SD/－Trp 培养盘，即每种盘涂布 3 个浓度（此步可放在铺完文库交配混合物之后再做）。

8. 正式铺文库交配混合物。将剩余的交配混合物全部铺于大约近 60 个直径 150 mm 的培养盘（根据预先确定的严谨筛选方案所制备的 QDO 或/和 TDO 盘），每盘铺 200 μL，相当于每

培养盘有 $2\times10^4\sim2\times10^5$ 细胞。为使二倍体细胞在盘上均匀分布，放 4～6 颗直径 5 mm 的灭菌玻璃珠并水平晃动以促进细胞的均匀分布，直至所有液体都被吸收。

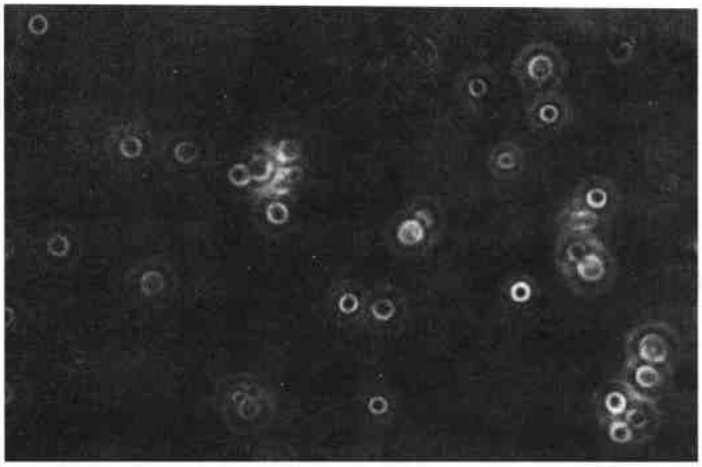


图 8－8　交配产生的合子

各叶代表 2 个源自双亲的单倍体细胞和出芽的二倍体细胞

9. 于 30 ℃倒置培养直至克隆出现。一般 TDO 培养基需 3～8 d，QDO 培养基则需 8～21 d。2～3 d 后有些克隆即能在文库筛选盘见到，但培养盘应至少孵育 8 d(TDO)或 16 d(QDO) 以容许生长缓慢的克隆（如弱相互作用）出现。对那些小的、苍白的克隆（其可能在两天后出现，但直径永远不会超过 1 mm)可忽略，能真正表达 Ade＋，His＋基因的克隆非常大，直径可＞2 mm，呈白色或微粉红色，且在－Ade 的培养基中将慢慢变红。另外，Y187 生长较 AH109 或 Y187/AH109 二倍体慢，且形成的克隆也小。

10. 计数 SD/－Leu，SD/－Trp 和 SD/－Leu/－Trp 盘上的克隆数。计算酵母交配的效率和筛选的克隆数。

11. 在 8～21 d 的酵母生长过程中，逐步将阳性克隆挑至 QDO 网格盘中，此时应开始在网格上对克隆从 1，2，3…开始逐一编号。4 ℃保存。依据克隆数多少，决定做 β－gal 检测的时间。网格盘是指使用直径 90 mm 的培养皿，在其底部用记号笔划成 $1mm^2$ 左右的网格，每格挑入一个阳性克隆。

由于在氮源受限时，二倍体细胞有进行减数分裂(meiosis)和形成孢子的潜在可能，所以文库筛选中所获得的阳性克隆不能放置在同一块培养盘上超过 2～4 周时间。否则，应进行冻存(置于－70 ℃)。总之，4 ℃保存时间不要过长。

12. 对生长在 TDO 基质中的克隆，将其复制到含 X－α－Gal 的 QDO 基质，并在 30 ℃孵育 3～8 d。

无论在 QDO 或 TDO 培养基中生长的转化子都是真正的双杂合阳性克隆，其中有些也许会包含多个 AD 文库质粒。然而最常见的假阳性克隆可通过筛选 *ADE*2，*HIS*3，*MEL*1 这 3 种报告基因的表达而消除。其他类型的假阳性可按照第五节中的方法来消除。

(二) 筛选文库的同时需做的检测

1. 文库滴定。

2.计算交配效率与筛选的文库克隆数。文库筛选前,需验证表型和检查对照菌的交配效率。由于交配效率在各次实验中不同,而交配效率又会影响所筛选出的文库克隆的数目,所以每次文库交配实验时都应平行设立对照交配实验。

三、文库滴定(Library Titering)

(一) 要点

1.稀释的文库不如未稀释的文库稳定,因此一旦文库被稀释,应在 1 h 内用完,否则滴度将显著降低。

2.培养基应使用推荐浓度的抗生素以抑制污染细菌的生长。

(二) 试剂和材料

1. YPDA/Kan 培养液(注意:需加抗生素)。

2. SD/ - Leu 培养盘。

(三) 步骤

1. 将 5 μL 文库转至有 1 mL YPDA/Kan 的 EP 管中,轻轻振摇混匀,此即稀释液 A(稀释系数 = 10^{-2})。

2. 从稀释液 A 中移 10 μL 至有 1 mL YPDA/Kan 的 1.5 mL EP 管中,轻轻振摇混匀,此即稀释液 B(稀释因子 = 10^{-4})。

3. 从稀释液 B 中各取 50 μL、100 μL,如上述涂布于 SD/ - Leu 盘。

4. 30 ℃,倒置培养 3 ~ 5 d。

5. 计算培养盘中的克隆数(应使用含 30 ~ 300 个克隆的培养盘)。

6. 依如下方法计算滴度(cfu/mL,克隆形成单位/每毫升交配液)。

公式:克隆数/铺盘体积(mL) × 稀释系数 = cfu/mL

举例:培养盘克隆数 = 400 ~ 600

铺盘体积 = 0.05 mL

稀释倍数 = 10^{-4}

滴度为:$400/0.05 \times 10^{-4} - 600/0.05 \times 10^{-4} = 8 \times 10^{7} - 1.2 \times 10^{8}$ cfu/mL

注意:应保证移液器的计量准确和操作技术的准确,否则,滴度计算结果可相差 2 ~ 5 倍。

四、计算酵母交配的效率和筛选的克隆数

(一) 交配效率

计数克隆约在 30 ~ 300 之间的 SD/ - Leu、SD/ - Leu/ - Trp、SD/ - Trp 培养盘中的克隆数

举例:SD/ - Leu 培养盘克隆数为 700。

SD/ - Leu/ - Trp 培养盘克隆数为 50。

交配效率 = 50/700 × 100% = 7%。

(二) 估算筛选的克隆数

公式:交配效率 × 文库滴度 × 总铺盘体积。

举例:$7\% \times (8 \times 10^{7} - 1.2 \times 10^{8}$ cfu/mL) × 12.5 mL = $7 \times 10^{7} - 1.05 \times 10^{8}$。

注:12.5 mL 总铺盘体积来自"筛选步骤中的 6",终止交配后,经两次冲洗,最后用 10 mL

0.5× YPDA/kan 悬浮细胞团块，测量细胞和培养基的总体积为 12.5 mL。尽管细胞体积可能会略有差异，但交配混和物都会达 12 mL，也正是如此，所以事先应制备培养盘 60～63 个。

第五节　对初步获取的阳性克隆的分析与验证

对初步获取的大量阳性克隆，要经过一系列的去伪存真的筛选过程，最终方可得到能够与诱饵蛋白发生相互作用的蛋白。主要包括：克隆表型验证、β－半乳糖苷酶(β－galactosidase)分析(简称 β－gal 分析)、PCR 扩增 AD/文库质粒、酶切归类、酵母体内、外再验证和测序。

一、克隆表型验证——大规模筛选

1. 当克隆生长至 21 d 时，应从原始交配盘中分别将所有克隆挑入 QDO 网格盘中，并逐一编号，过小的克隆应摒弃。

2. 在滤纸上进行 β－gal 分析(实验步骤见下)。注意：蓝色克隆为阳性。

3. 划线筛选。利用 β－gal 分析后的同一网格盘，若盘上的克隆未全部被滤纸所“蘸”掉，挑取每个克隆的剩余部分在新的 QDO 盘上划线。若所剩克隆体积过小，则需在 30 ℃内培养 2～4 d 后再划线。为节省培养盘，可在其底部用记号笔划线分割成 4～8 部分。此时必须严格按照网格盘上的阳性克隆号标在划线盘上，做到对号入座。30 ℃，倒置培养 4～6 d。

4. 再次划线筛选。从第一次划线盘上挑取单克隆在新的 QDO 盘上再次划线。如果第一次划线产生的克隆体积很小，或未能很好地分离，则应挑取“一堆”克隆划线。30 ℃，倒置培养 4～6 d。

5. 从第二次划线盘上挑取单克隆入新的 QDO 网格盘中(注意编号书写准确)。第一二次划线盘暂时存放于室温。

6. 再次 β－gal 分析。凡蓝色克隆均可初步认定是阳性克隆。

7. 对网格盘上的每个阳性克隆进行保种，同时进行酵母质粒的提取。由此可见，克隆的大小极为重要，划线盘暂时存放的目的也在于此。

注意：两次划线之目的，在于随着选择性压力的维持状态，使没有发生相互作用的质粒被丢掉，同时也使真正的阳性克隆维持其正确的表型。每次经 β－gal 分析和划线筛选后，克隆数目都将逐渐减少，因此每次准确标记克隆的编号是一非常重要的工作。

二、β－半乳糖苷酶活性分析

(一) 一般概念

β－半乳糖苷酶显色反应选择法是利用表型特征筛选(又名遗传检测法)、检测 *LacZ* 基因表达的一种常用的经典的方法，其原理在很多书中已有介绍。在酵母双杂交系统中主要用它分析两个已知蛋白的相互作用。该方法以 X－gal 作为 β－半乳糖苷酶的底物，分固相和液相两种分析方式，前者又分为琼脂板分析和滤纸分析。

1. 琼脂板分析是一种定性分析，指在制备 SD 选择性培养盘时，培养基内含有 X－gal，但其敏感度较低，一般不推荐在酵母双杂交中使用。

2. 滤纸分析也是一种定性分析，是指在不含有 X－gal 的 SD 选择性培养盘的克隆表面覆盖滤纸，以滤纸“蘸取”克隆，对滤纸上的克隆进行显色反应，以蓝色克隆确定为阳性。此方法方便、敏感、背景低、价格便宜，适合于大规模筛选。

3. 液相分析是定性分析，虽然其产生的背景低，但其敏感性也低，不能对反应弱的克隆给予定性，且不适宜做大规模筛选。本章仅就滤纸分析法予以介绍。

（二）材料准备

1. Whatman #5 或 VWR Grade 410 滤纸。将滤纸剪成与网格盘同样大小形状，可留有两个小把手以利于提取，高压灭菌，按每个网格培养盘使用一张滤纸准备，同时还应准备相应数量的直径大于网格培养盘 2～4 cm 的滤纸。

2. 镊子（2～4 把）、牙签、针头、直径 120 mm 的培养皿一个，所有物品需灭菌。

3. Z 缓冲液：

$Na_2HPO_4.7H_2O$	16.1 g/L
$NaH_2PO_4.H_2O$	5.50 g/L
KCl	0.75 g/L
$MgSO_4.7H_2O$	0.246 g/L

调 pH 值至 7.0，灭菌，室温可放 1 年。

4. X－gal 储备液：X－gal 即 X－β－gal，也即 X－β－半乳糖苷，全称为 5－bromo－4－chloro－3－indolyl－β－D－ galactoside。溶解 X－gal 于二甲基甲胺（N，N－ dimethyl formamide，DMF）中，终浓度为 20mg/mL，避光贮于－20 ℃。

5. Z 缓冲液 /X－gal 混合液（现用现配）：

Z 缓冲液	100 mL
β－巯基乙醇（β－ mercaptoethanol）	0.27 mL
X－gal 储备液	1.67 mL

6. 用实验室常用的保温瓶倒入液氮。

（三）实验步骤

1. 为求最佳结果，用新鲜（如 30 ℃生长 2～4 d）的，直径在 1～3 mm 的克隆。若仅有少量克隆需分析，将其直接划线或取小堆克隆铺于相应的 SD 选择琼脂盘，于 30 ℃再孵育 1～2 d，随后用下述 β－半乳糖苷酶分析。此处是指对从原始交配盘中挑入 QDO 网格盘的克隆检测。

2. 将 Z 缓冲液/X－gal 溶液，放在 120 mm 的培养皿中。

3. 将一张直径大的滤纸放入盛有 2.5～5 mL Z 缓冲液 /X－gal 混合液的培养皿中浸湿，取出后放在托盘上。

4. 用镊子将一灭菌的干滤纸覆盖于培养盘的克隆面，与克隆紧密接触。一旦覆盖，不可再移动滤纸。轻轻用镊子摩擦滤纸尽量使克隆粘至滤纸上。

5. 用针头经滤纸刺 3 个或更多不对称的洞至琼脂以定位。

6. 用镊子小心地将滤纸移离琼脂盘。若滤纸上的克隆过少时，以牙签挑取相应位置的克隆至滤纸，应特别注意克隆在滤纸与琼脂盘上的对称位置，不可弄错。将滤纸的克隆面朝上，放入液氮中，浸没 10 s，使克隆固化于滤纸上。

7. 将滤纸移出液氮，平衡至室温（冻/融处理可快速、有效地裂解细胞）。

8. 小心将拓有克隆的滤纸（克隆面朝上）放在托盘中预浸湿的滤纸上，滤纸下面及二层滤纸间避免气泡。最后将托盘加盖，平放于室温。

9. 30 ℃（或室温）孵育滤纸并定时观察蓝色克隆出现的时间。

（四）几个值得商榷的问题

1. 各个克隆产生 β－半乳糖苷酶蓝色显色反应的时间不同，文库筛选中一般进行 30 min 至 8 h，延长孵育（＞8 h）可能会产生假阳性。

2. 用 β－半乳糖苷酶阳性对照质粒转化的酵母在 20～30 min 内变蓝。CLONTECH 公司的 *Yeast Protocol Handbook* 中指出，大部分双杂交阳性对照的共转化菌株在 60 min 内出现阳性蓝色信号，但笔者在操作中发现，一般在 1.5 h 后才开始出现蓝色反应。

3. 初次筛选的克隆数依不同的实验可能有很大差异，笔者所在实验室曾在初次筛选中获得近 2 000 个克隆，这其中肯定会有相当数量的假阳性，且如此筛选下去，工作量也将极大。于是在 β－gal 分析过程中采取分时间段标记阳性克隆的方法，即每 2～2.5 h 对出现的克隆标记 1 次，共标记 3 次（可用不同颜色的记号笔以区分之）。最后，取各时间段部分克隆，向下进行实验。此种作法可以人为控制所需的克隆数量，但也可能丢掉部分信息，此方法仅供参考。

总之，β－半乳糖苷酶滤纸分析是采用拓字法将克隆移至滤纸上，通过冻融法裂解酵母细胞壁，将下层滤纸预浸湿于 Z 缓冲液/X－gal 溶液中，利用虹吸作用使上层滤纸发生显色反应的一种筛选法。

三、酵母质粒 DNA 的提取

将第 2 次 β－半乳糖苷酶分析使用的网格盘，再放 30 ℃生长 1～2 d 后，对阳性克隆进行保种和酵母质粒 DNA 提取。若阳性克隆依旧很小，则可从第 2 次划线盘上挑取克隆，但尽量使用网格盘，以减少假阳性的几率。

（一）材料准备：

1. 含目的质粒的酵母菌落（网格盘或划线盘）。

2. SD/－Ade/－His/－Trp/－Leu 液体培养基（若用 TDO 盘筛选，此处乃使用 TDO 液体培养基）。

3. 酵母裂解缓冲液：

	终浓度
Triton X－100[V/V]	2%
SDS[W/V]	1%
NaCl	10 mM
Tris·HCl	10 mM
EDTA	1 mM

4. 酚：氯仿：异戊醇 = 25:24:1(V/V/V)（用缓冲液平衡酚）。

5. 3 M 醋酸钠（NaAc）pH 5.2。

6. 直径 0.4～0.5 mm 的酸洗玻璃珠（Sigma 公司），可直接使用。

（二）操作步骤

1. 挑取单克隆，投入放有 1 mL SD/－Ade/－His/－Trp/－Leu 液体培养基的 Eppendorf 管中，振荡混匀，移至预先放有 4 mL 相同液体培养基的摇菌管中。

2. 30 ℃ 230 r/min，振荡培养 24～36 h 至细胞静止期。

3. 使用 Eppendorf 管收集细胞，1 000 g，室温离心 5 min，弃上清。

4. 加 200 μL 酵母裂解液，0.2 g 酸洗玻璃珠（约 200 μL 体积）及 200 μL 酚：氯仿：异戊醇，

剧烈振荡 2 min。

5. 12 000 r/min,室温离心 5 min。

6. 将上清转入另一新的 Eppendorf 管,加入 0.1 倍体积的 3M NaAc(pH5.2)和 2.5 倍体积无水乙醇,混匀,-20 ℃放置至少 30 min。

7. 12 000 r/min,4 ℃,离心 5 min,弃上清。

8. 用 70%乙醇清洗沉淀,至少两次,以确保将盐离子洗掉,这对下一步的电转极为重要。

9. 室温干燥,用 10~15 μL 去离子水溶解。

注意:从每个阳性酵母克隆中分离的质粒 DNA 是 DNA-BD/bait 质粒和至少一种 AD/library 质粒的混合物。酵母质粒的提取步骤并不困难,但产量较少,因此,要用小量体积的水溶解。

四、酵母质粒转入大肠杆菌 DH5α

(一) 材料准备

1. LB 细菌培养液 500 mL(使用 1 L 容积的三角瓶)。

2. LB/氨苄培养盘。

3. WB 溶液 2 L(10%超纯甘油,90%去离子水,V/V),需高压灭菌。

4. 其他:液氮、50 mL 离心管、Eppendorf 管、摇菌管(用于提细菌 DNA)。

5. 电转仪(gene pluser),电转杯(genc pluser cuvette)均为 Bio Rad 产品。

(二) DH 5α 感受态细菌的制备

1. 挑取细菌单克隆于含有 10 mL LB(无抗生素)的 50 mL 离心管中,37 ℃,180 r/min 振荡 14 h。

2. 取 0.8 mL 菌液加入 500 mL LB(无抗生素)中,37 ℃,170 r/min 通气振荡 6~7 h(最多不超过 8 h),直至 OD_{550} = 0.8。

3. 收集细菌,冰浴 40 min。

4. 于 4 ℃,3 000 r/min 离心 10 min,沉淀细菌。

5. 将细菌沉淀悬浮于 250 mL 无菌的、冰预冷的 WB 溶液。

6. 于 4 ℃,3 000 r/min 离心 10 min。

7. 重复步骤 5、6 两次。

8. 轻轻倒掉上清,小心吸去大部分上清,剩约 5 mL。将细菌悬液转移至 50 mL 离心管。

9. 4 ℃,3 000 r/min 离心 10 min,小心弃上清,剩约 1 mL 菌体。

10. 加 1~1.5 mL WB 溶液悬浮细菌,使其浓度为 3×10^{10}细胞/mL,并分装于用液氮预冷(非常重要)的 Eppendorf 管中,每管约 25 μL(视样品多少作相应量的感受态细菌,并应保证每管内的细菌浓度)。

11. 放入冰中,立即做电转。

注意:刚开始收集的细菌,离心后呈淡黄色至微绿色,且不易悬浮,经 WB 溶液的清洗,菌体变白,并随上清的倾到,极易流失。

(三) 电转化

1. 取 1 μL 酵母质粒入 25 μL 预冷的感受态细菌中,混匀。

2. 转入到电转杯中,避免气泡。

3. 上机。电转参数：电压 2 500 V；电阻 200 Ω；电容 25 μF。电转时间约持续 4 s。

4. 电转杯中加 1 mL LB，充分混均，移回放感受态细菌的 Eppendorf 管中。

5. 37 ℃，250 r/min 振摇 45 min。

6. 2 500 r/min，4 ℃ 离心 1～2 min。

7. 吸弃 850 μL 上清，将 150 μL 混匀，涂布于 LB/氨苄盘。

8. 37 ℃倒置培养，不能超过 20 h（若超过 20 h 的菌落为“卫星”菌落，假阳性居多，视“不可信菌落”，不必摇菌。）

（四）细菌质粒 DNA 的提取

同一般质粒 DNA 的提取。同时冻存菌种。

（五）注意与说明

1. 感受态细菌、电转杯始终放入冰中。电转应在低温下（0～4 ℃）进行，如果在室温下操作，转化效率可能会降低到原来的 1/100。

2. 电转化时如果酵母质粒洗涤不干净，有各种离子的存在（主指盐离子），将造成电转杯的击穿甚至更严重的事故。

3. 电转杯的清洗：自来水、去离子水各洗 3 次，70% 乙醇和无水乙醇各浸泡 10 min，自然晾干。

4. 制备感受态细菌时摇菌的 LB 不加氨苄青霉素。

5. 获得的阳性克隆可视为文库质粒，而文库质粒的细菌筛选表型为氨苄抗性，故电转化子应铺于 LB/氨苄盘，以选择性扩增 AD 质粒。

6. 除电转化入大肠杆菌 DH 5α 外，还可使用其他菌株，如 HB101 等。

五、克隆归类

初步筛选所得到的阳性克隆，实际上是 DNA－BD/bait 质粒和至少 1 种 AD/文库质粒相互作用的混合物。通过两次划线、β－半乳糖苷酶分析以及在氨苄盘中筛选转化子等若干步骤的表型再验证后，逐步使二倍体酵母体内的非相互作用质粒在选择性压力下丢失，以得到那些在交配中真正与诱饵质粒发生相互作用的文库质粒。即使这样，由于一个酵母体中可以含有多个质粒，且插入在 AD/文库质粒中的许多 cDNA 序列是相同的、近似的或重复的，仍会对以后的工作造成很多麻烦。如何识别、删除这些重复序列？设计者采用 PCR 扩增与酶切相结合的方法。先通过 PCR 扩增得到 AD/文库质粒中的插入子，再对扩增片段进行酶切、归类，达到纯化、精简阳性克隆种类与数量的目的。

酵母双杂交系统中，cDNA 文库插入在 AD/文库质粒中。为保证对这些插入子的扩增效率，CLONTECH 公司的 MATCHMAKER Libraries 试剂盒，已在构建质粒时设计了扩增插入序列的 PCR 引物（图 5）。引物可直接购得或合成。经 PCR 扩增得到欲筛选的片段，其大小是不一致的，从几百 bp 至几个 kb。可先根据 PCR 扩增片段的大小，做一粗略的分类，再对每类中的全部或部分 PCR 结果用限制性内切酶消化。

CLONTECH 公司推荐使用 *Alu* I 或 *Hae* III 两种酶并分别进行单酶切，此二酶不像一般的内切酶可识别 6 个核苷酸序列，它们仅可识别 4 个核苷酸序列。前者识别序列 AGCT，后者识别序列 GGCC。理论上，4 个核苷酸序列出现的频率大于 6 个核苷酸序列，也就是说，在所得到的某一序列中，由于这 2 种酶的酶切位点多于其他酶，可将这一大片段切割成几个片段。即使

扩增片段大小相同，也有可能产生不同条带。利用这一特点，对每个 PCR 产物进行酶切，每个单酶切都会产生几个条带，很少出现单一条带，多者可达 5～6 条。于是可将电泳条带大小相同或相近的克隆归为一类，再从每一类中选取“代表”测序。

关于 PCR 扩增、酶切、测序及序列分析等，参照一般分子生物学技术指导。分析归类则要根据不同实验自行设计。此处虽叙述简单，但实际仍包含大量工作，且最终结果要由此获得。

六、在酵母体内再验证蛋白的相互作用

为进一步确证经上述筛选得到的和诱饵蛋白发生相互作用的蛋白是否真实，将两种质粒同时转入一个宿主菌以再次验证。可选择如下实验方案：

（一）共转化

1. 制备 AH109 感受态酵母菌。

2. 从细菌中提取的文库质粒 DNA 逐一和 DNA－BD/bait 质粒做共转化。

3. 铺于 SD/－Ade/－Hie/－Leu/－Trp/X－α－gal 盘。

4. 30 ℃孵育直至克隆出现。

（二）酵母交配试验

酵母交配试验是将两种质粒转入同一宿主细胞中。在有多个表达 Ade^+，His^+，Mel^+/$LacZ^+$ 阳性克隆待分析的情况下，这种方法更为可信。

1. 分别制备 AH109 和 Y187 感受态酵母菌。

2. 将所有获得的 DNA－AD/文库质粒逐一转化入 AH109，在 SD/－Leu 培养盘中选择。

3. 将研究者构建的 DNA－BD/bait 质粒和阴性对照 DNA－BD 质粒 pGBKT7－Lam 分别转化入 Y187，在 SD/－Trp 培养盘中选择。

4. 制备 SD 培养盘（非缺陷型），并划有网格线。

5. 逐一挑取 AH109[DNA－AD/library]和 Y187[DNA－BD/bait]克隆，在 SD 网格盘上交配。

6. 再逐一挑取相同编号的 AH109[DNA－AD/library]克隆和 Y187[DNA－BD/pGBKT7－Lam]克隆在网格盘上交配。交配方法见第四节二倍体细胞的交配产生。

7. 30 ℃，交配 4 h。

8. 取 AH109[DNA－AD/library]×Y187[DNA－BD/bait]交配混合物划线于 SD/－Ade/－His/－Trp/－Leu 盘。

9. 取 AH109[DNA－AD/library]×Y187[DNA－BD/pGBKT7－Lam]交配混合物划线于 SD/－Trp/－Leu 盘。

10. 30 ℃，倒置生长 4 d。

11. 分别挑取生长在 SD/－Ade/－His/－Trp/－Leu 盘上的单克隆至另一相同培养基的网格盘上（阳性结果盘）。

12. 再分别挑取生长在 SD/－Trp/－Leu 盘上的单克隆至另一相同培养基的网格盘上（阴性对照盘）。

13. 分别对阳性结果盘和阴性对照盘做 β－半乳糖苷酶活性分析。

14. 观察结果与分析。理论上，阳性克隆经多次筛选，基本均为蓝色反应，而阴性对照盘均应为白色克隆。就一个克隆而言，若在阳性盘上为白色，不论阴性盘颜色如何，应摒弃；若在阳性、阴性盘上均为蓝色，也摒弃；必须在阳性盘上为蓝色，同时阴性盘上为白色的克隆方为真

正的阳性结果。

15. 对阳性克隆保种冻存。

注意：①酵母交配实验中，待检测的 DNA－AD/文库质粒是转化 AH109，但仍涂布于 SD/－Leu 培养盘；②因所涉及的步骤较多，时刻注意克隆编号的正确性，切勿混乱；③本实验所涉及的方法，前面均有介绍，此处未再叙述；④实验前一定要根据待测阳性克隆的数目及步骤，做好制备培养盘种类与数量的计划，以防浪费或不足，其他质粒所需培养盘相对容易准备；⑤酵母体内验证和 PCR 可同时进行。

七、在酵母体外分析和验证蛋白的相互作用

（一）体外转录、翻译融合蛋白

CLONTECH 公司的 DNA－AD 文库质粒本身已含有 T7 启动子和 HA 表位标签，利用公司提供的 MATCHMAKER Co－IP Kit，可以直接在体外进行融合蛋白的转录翻译。

（二）免疫共沉淀

一个比较常用而公认的在酵母双杂交后检测相互作用蛋白质的方法。笔者实验室曾用该方法取得过肯定的结果。因该方法已不属于酵母双杂交实验技术范围，故不作介绍。

八、测　　序

对经酵母体内再验证和酶切归类后的全部阳性克隆进行测序（注意：应使用从细菌中分离的质粒），最终得到的即是和诱饵蛋白相互作用的蛋白质。

第六节　酵母双杂交技术的应用

一、酵母双杂交技术的应用

蛋白－蛋白交互作用（protein－protein interaction）是生命活动持续的最基本的根据。酵母双杂交系统的建立为研究这一问题提供了有力的手段，其主要应用于以下几个方面。

（一）研究已知蛋白间的相互作用

对于一些经遗传学、生物化学、生理学、分子生物学等实验提供的可能存在相互作用的蛋白，可以用酵母双杂交分析做进一步验证。因为双杂交系统的灵敏度高，而且反应了蛋白在生物体内的交互作用。例如，癌蛋白 Jun 与 Fos 之间的相互作用，抗癌蛋白 *P*53 自身间的相互作用，以及 Ras、Raf 的交互作用等都是应用酵母双杂交方法确定的。

（二）寻找在蛋白－蛋白相互作用中起关键作用的结构域

在证实两个蛋白能发生交互作用后，可以对该蛋白质进行缺失实验。通过缺失掉不同的片段来验证该蛋白在双杂交系统中是否仍具有激活转录的功能，从而对蛋白质中起作用的功能域进行精确的定位。

（三）寻找与靶蛋白相互作用的新蛋白

这是研究者最感兴趣的问题，也是双杂交系统最广泛、最有价值的应用。本章所提供的方法即是基于这一目的的应用。但双杂交仅提供蛋白质间能发生作用的可能性，这种可能还必须经过其他实验验证，尤其要与生理功能的研究相结合。

二、酵母双杂交技术的缺陷

尽管该系统已被证实为一种非常有效的方法，但它也有自身的缺点和问题。①它并非对所有蛋白质都适用，这是由其原理所决定的。双杂交系统要求两种杂交体蛋白都是融合蛋白，都必须能进入细胞核内。因为融合蛋白相互作用激活报告基因转录是在细胞核内发生的。②假阳性的发生较为频繁。所谓假阳性，即指未能与诱饵蛋白发生作用而被误认为是阳性反应的蛋白。而且部分假阳性原因不清，可能与酵母中其他蛋白质的作用有关。③在酵母菌株中大量表达外源蛋白将产生毒性作用，从而影响菌株生长和报告基因的表达。

三、使用酵母双杂交技术应注意的问题

真正明了酵母双杂交技术的主要原理及筛选方法是进行酵母双杂交实验的前提，构建成功的诱饵质粒及大量的材料准备是进行酵母双杂交实验的保证。只有明了双杂交的原理，才有可能设计实验进程、才能有目的的进行材料准备，并能对实验结果作出预测与分析，尤其要对具体实验中各种选择性压力培养基的使用目的要十分清楚。大量的材料准备、较长的实验流程是酵母双杂交有别于其他实验的特点，而其操作技术本身并不十分困难。特别应提出的是，一个阳性克隆的编号往往要被反复记录多次，因此，要时时注意编号的正确性。另外，若从公司购得待筛选的酵母 cDNA 文库，应注意不同的公司有不同的产品，且各公司的产品不断更新换代，要认真阅读实验指导手册，以防出现失误。

相信随着分子生物学技术的发展与推广，酵母双杂交技术在今后的蛋白质组学研究中将发挥更大的作用。

（刘冬青　孟宪敏）

第九章

SELEX 技术

内容提要

SELEX 技术是一种新的组合化学技术。本章阐述了该技术的基本原理、随机寡核苷酸库的修饰与优化、影响因素等；指出了评价寡核苷酸配基结合特异性的指标及适配分子在各方面的应用；简单介绍了 SELEX 的技术改进与发展。

生物大分子的定性和定量检测在基础研究和临床实践中具有极重要的作用，尤其是随着人们对生物反应的高度特异性和高效性的认识，对检测方法分辨率和灵敏度的要求也越来越高。为适应新的检测需要，人们在对现有的方法进行改进的同时，发展了许多新的、特异的、精确的检测方法，出现了许多挑战传统的新方法。

今天，技术的发展已经使人们能够根据自己的需要，改变生物分子的特性，得到自然界不存在的物质。这与仅利用天然存在的分子来进行分析，其前景是截然不同的。适配分子技术就是这样一种最近发展起来的体外筛选和放大技术，通过该技术可得到一段能与各种目标分子高亲和、高特异结合的核酸序列，该核酸序列称之为适配分子(aptamer)。Aptamer 一词来源于拉丁文 Aptus，意即适合的意思(fit)，而这种体外的筛选技术称之为指数级富集的配体系统进化技术(systematic evolution of ligands by exponential enrichment, SELEX)。简言之，Aptamer 就是一段核酸序列，它能与相应的配体高特异性和高亲合力地结合。从原始文库中筛选配体相应的 Aptamer 的技术称为 SELEX 技术。Aptamer 具有广泛的实用性，它能用于任何以分子识别为基础的诊断和治疗。在这一概念提出仅 8 年之后，Aptamer 的临床治疗应用已取得重要进展，同时 Aptamer 也被用于体内诊断和体内成像。

第一节　SELEX 技术简介

SELEX(systematic evolution of ligands by exponential enrichment)，中文名称为指数级富集的配体系统进化，是一种新的组合化学技术。它是应用大容量的随机寡核苷酸库，并结合 PCR 体外扩增技术，以指数级富集与靶分子特异结合的寡核苷酸，经过几轮或数十轮筛选过程，获得高亲和力、高特异性的核酸配基(aptamer)。自 20 世纪 90 年代初 Tuerk 等(1990)和 Ellington 等(1990)首次利用该技术成功地筛选出 T4DNA 聚合酶和有机染料的 RNA 特异配基以来。应用 SELEX 技术筛选过的靶分子已从开始的一些核酸结合蛋白，例如 DNA 聚合酶、复制酶、逆转录

酶等，扩展到非核酸结合蛋白如细胞生长因子、激素、小肽、抗体以及 ATP、氨基酸等小分子有机物，甚至金属离子。与此同时，SELEX 技术本身也得到不断发展，许多改良的 SELEX 技术如导向 SELEX(blended SELEX)、基因组 SELEX(genomic SELEX)、复合靶分子 SELEX(complex target SELEX)等相继建立。目前，人们认为，对于任何一种蛋白质，都可以利用 SELEX 技术寻找到一个可与之相匹配的寡核苷酸配基。这种寡核苷酸配基与靶分子结合的解离常数(K_d)可达 $10^{-9} \sim 10^{-12}$ mol/L 水平，能分辨同源蛋白或几乎完全一样的小分子量化合物，亲和力和特异性往往高于其他任何类型的配基，因此 SELEX 技术已经在基础医学、临床诊断及药物筛选方面得到广泛应用。

第二节　SELEX 技术的基本原理

一、SELEX 技术的基本路线

从大量的随机序列的寡核苷酸库中鉴定出一种数量很少的具有独特性质的核酸序列的 SELEX 方法是由 Tuerk 和 Ellington 等在 1990 年建立的，它是一种通过体外反复选择和放大从巨大的核苷酸组合库中筛选特定核苷酸序列的方法。组合库一般是由重复单体生物多聚物组成，如核酸、肽，这类分子很容易地被逐级放大，使筛选过程变得快速而容易。

SELEX 过程，首先是用组合化学合成 DNA 的方法合成随机序列的核酸库，库中的每一个成员是一个线形的寡聚物，库中分子多样性的程度取决于随机寡核苷酸的长度。理论上讲，一个 40 个碱基的随机序列可有 1.2×10^{24}种核苷酸排列顺序($4^{40} = 1.2 \times 10^{24}$)。实际上 1 μmol 的固相 DNA 合成的随机组合核苷酸库的序列多样性仅限制于 $10^{14} \sim 10^{15}$种，能否找出与靶分子相互作用的稀有序列，很大程度上取决于组合库的多样性，在所有种类的组合随机序列库中寡核苷酸库的多样性最为丰富。

筛选过程中，首先在一定的温度下将寡核苷酸随机序列库与靶分子共同孵育于特定的缓冲液中，组合库中的极少数分子将与靶分子结合，然后用物理的方法将结合的分子与未结合的分子分离开来。标准的方法是将蛋白靶分子固定于硝酸纤维素滤膜上，而其他小的靶分子则固定于固相载体的表面，因此，未与靶分子结合的序列很容易被冲洗掉，而与靶分子结合的序列将被分离出来并放大以获得一个序列库，再对这一序列库进行下一轮的筛选和放大。每一轮选择的严格性决定了高亲和结合子的增强效率。高亲和结合的增强进度可通过与靶分子的结合分析来确定。经过数轮的筛选和放大后，一旦达到亲和饱和，所得到的增强库即被克隆和测序，以获得筛选库中每一个序列的信息。单个序列的特异性将根据其与靶分子结合的能力做进一步的定性。通常情况下，大多数序列(90%以上)与靶分子有很好的高选择性亲和力。通过 SELEX 筛选出来的 Aptamer 是含有一段固定序列的全长序列，这个固定序列主要用于协助放大过程，在放大过程中起到引物的作用。这种全长 Aptamer 一般由 70～80 个碱基构成，通过剪切可去除那些对结合靶分子无关紧要的片段，或折叠成最利于结合靶分子的结构。剪切 Aptamer 主要是用于确定最小的靶分子结合域，目前已得到数种少于 40 个核苷酸的功能性 Aptamer。在大多数情况下，Aptamer 的固定片段主要用于放大，而对 Aptamer 的功能(即识别功能)无关紧要，因此可以去除该段的固定片段。现已发现了一种无须固定引物序列的 SELEX 方法，从而筛选出较短的 Aptamer。

SELEX 技术之所以能够建立，是由于形成寡核苷酸的碱基之间的相互作用往往形成许多

空间结构，如发卡（hairpin）、假结（pseudoknot）或凸环（bulge）等，这些结构非常容易与各类靶分子结合。一个足够容量的寡核苷酸库包容了所有可能的立体结构，在与靶分子结合时，几乎不需要失去"熵"就能存在一个适配的结构，因此有可能从中筛到任何种类的靶分子的高亲和力配基。

Aptamer 鉴定过程（SELEX）所需循环的次数取决于每一轮筛选放大的严格程度及靶分子的特性。对大多数靶分子而言，经过 8～15 次循环即可达到较强的亲和力。一般研究人员能在 2 天内完成一次循环，包括克隆和测序，一次典型的 SELEX 过程需要 2～3 月。一旦序列被确定，即可用化学合成制备 Aptamer。小规模的合成与纯化可在 3 天内完成，得到 nmoles 级的 Aptamer，用于诊断分析的设计和实验。整个过程比制备单克隆抗体所需的时间短得多。最近有人研究出了一种进行 SELEX 过程的自动平台，能更快、更经济地鉴定和筛选 Aptamer。该自动平台能在数小时内完成一次选择和放大循环，几乎不要人为干预。此平台能在微孔板上同时进行多个 Aptamer 的鉴定，大大提高了鉴定和筛选 Aptamer 的效率。

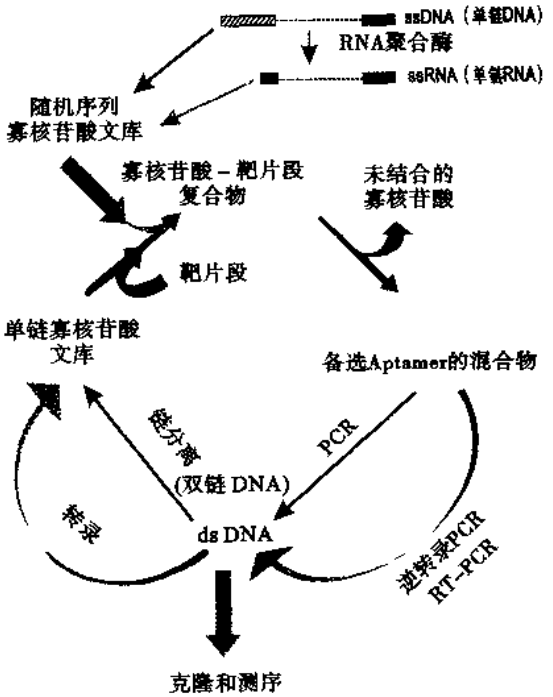


图 9－1　SELEX 技术的基本路线示意图

注：用组合化学的方法合成 ssDNA 随机序列库，一般有 10^{14}～10^{15}种序列。DNA 单链中间部分为随机序列，两端为固定的引物序列。中间随机序列区为 4 种三磷酸核苷的随机组合。为了获得非偏倚库，在合成库时，加入 4 种核苷酸的量相同。两端固定序列为引物结合位点。Aptamer 可为 DNA，亦可为 RNA。RNA 库可由 DNA 文库通过特定的酶反应产生，亦可由 RNA 聚合酶（如 T7 启动子聚合酶）在体外合成。寡核苷酸库与靶分子在特定的缓冲体系中孵育，库中少数分子与靶分子结合，形成复合物。用适当的分离技术，将结合序列与未结合序列分开。从复合物上分离出结合序列，形成含有高亲和力和低亲和力分子的寡核苷酸库。将获得的分离产物扩增，做进一步筛选，除去低亲和力分子。DNA Aptamer 可直接进行 PCR 扩增，RNA Aptamer 进行 RT－PCR 扩增。图 1 中黑箭头表示 DNA Aptamer 的筛选过程，灰箭头表示 RNA Aptamer 的筛选过程。从 PCR 产物分离出的单链 DNA 和靶分子孵育，进行下一轮筛选。通常需要数轮筛选，才能获得高亲和力的 Aptamer。当 Aptamer 达到亲和饱和时，对其进行克隆和测序。

Aptamer 以其高特异性见长。它能分辨靶分子上微小的结构差异，比如能区分靶分子上有

无甲基、羟基和手性等。Aptamer 有很高的特异性，有时其特异性甚至优于单克隆抗体。这是由于在 SELEX 过程中去除了那些能与类似物质结合的有关序列。实际上，一种叫做 Counter - SELEX 的方法能区别同类物质结构上的细微差别，在筛选过程中，可将那些与结构类似物有高亲和力的 Aptamer 有效地区分开来。当进行小分子分析诊断时，待分析的靶分子与其类似物都十分重要，有时不必去除这些类似物。

即使对某一种靶分子包括结构和特性在内的一切信息一无所知，但通过 Counter - SELEX 也能从复杂的混合物中找出这种独特的靶分子。例如可通过 SELEX 技术寻找一种存在于肿瘤细胞而不存在于正常细胞表面的独特表位，或者找出一种仅存在于病人血清中的分子标志物，在这一过程中可用肿瘤细胞的 Aptamer 扣除正常细胞的 Aptamer，或用病人血清 Aptamer 扣除正常血清 Aptamer 完成。

二、Aptamer 的修饰与优化

对 SELEX 筛选库的寡核苷酸进行化学修饰有利于 Aptamer 的应用，未修饰的寡核苷酸，特别是 RNA，通常很容易被体液中的核酸酶降解。寡核苷酸的修饰(图 9 - 2)已经取得了突出的进展，使寡核苷酸能抵抗核酸酶的作用。特别是对各种寡核苷酸骨架的修饰，取得了较好的核酸酶抗性。大多数骨架修饰是在化学合成的过程中完成的，不适于 SELEX 过程中的酶，但硫代磷酸修饰是一个例外。硫代磷酸寡核苷酸的连接处能与蛋白发生非特异结合，因此不适于 Aptamer 的筛选。某些糖骨架 2 位的修饰亦能产生核酸酶抗性。由于体液中最常见的核酸酶具有嘧啶特异性，因此嘧啶核苷 2 位上的替代能抵抗体液中核酸酶的降解。最重要的是，嘧啶核苷核糖 2 位被 NH_2 和 F 替代后，仍可作为 SELEX 过程中的酶底物，这样 Aptamer 可在体液中长时间存在，并能抵抗核酸酶的作用，故可广泛用于诊断和治疗。

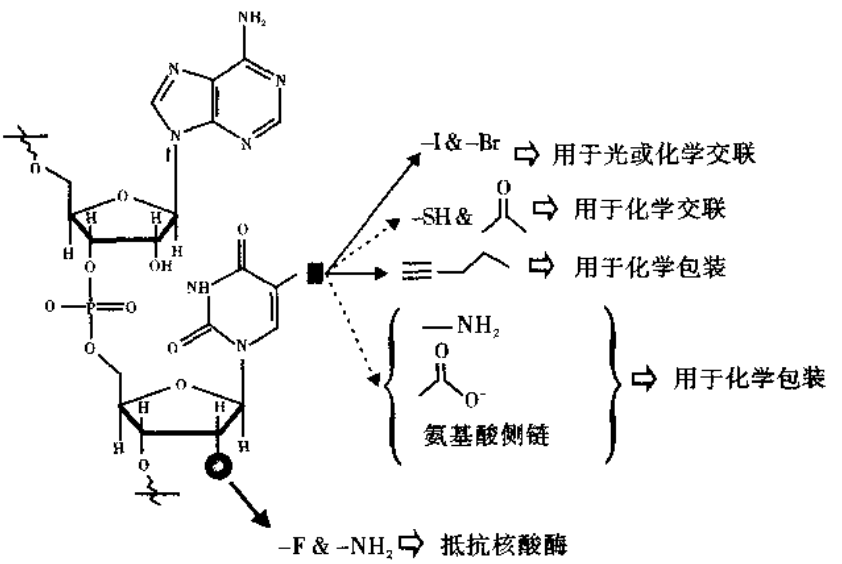


图 9 - 2　Aptamer 的修饰

注：戊糖 2′位修饰增加了核酸的稳定性，而嘧啶 C - 5 位的修饰可用来结合特定的靶分子或与靶分子发生共价交连。实箭头表明 SELEX 过程中已经用于 Aptamer 识别的修饰点，虚箭头表明潜在的 Aptamer 识别的修饰点。

在 SELEX 过程中仅有 4 种构建单体能形成核苷酸库。在 4 种单体核苷酸碱基上附加不同的替代基团可有效地增加核苷酸库的多样性。只有当修饰库能作为 SELEX 过程中的酶底物，这种修饰库才有意义。在这方面，已有数种非天然碱基(图 9－3)被成功地掺入 DNA 和 RNA。Piccirlli 和 Bain 等成功地将 Iso－G，Iso－C 及 2,6－双氨基嘧啶或黄嘌呤掺入 DNA 和 RNA；Tor 和 Dervan 等在体外转录过程中将甲基化嘌呤碱基掺入 RNA 中；其他附加物如生物素和 EDTA 等可在转录后加在 RNA 的碱性氨基上。此外核苷酸还可以尾标记的方法标上某些功能基团，如标上疏水基团有利于筛选疏水靶分子的 Aptamer。最近 Schweitzer 和 Guckian 等研制了一种无须氢键来识别的新碱基对，这种无极性、无氢键碱基能在酶的催化下掺入 DNA 中，该碱基类似物能作为核苷酸的构建单体，而与天然碱基的 Watson－Crick 碱基配对有所不同。上述修饰大大增加了核苷酸库的多样性，也增加了独特 Aptamer 的可能性。

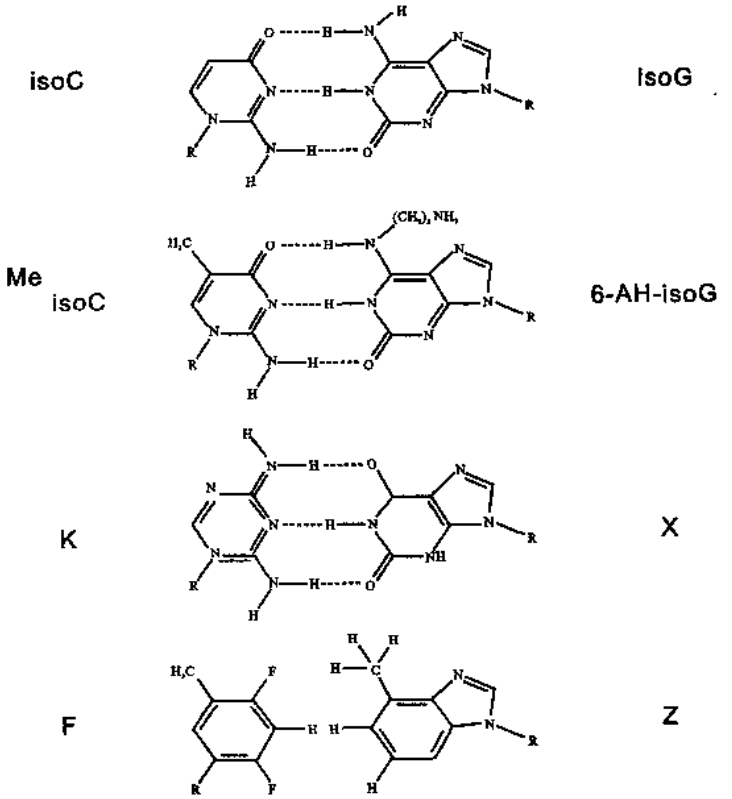


图 9－3　可被多聚酶聚合的非天然碱基

与含 2′－修饰的嘧啶寡核苷酸一样，含 5′－修饰的嘧啶寡核苷酸同样具有核酸酶抗性，如从胸腺嘧啶修饰的 RNA 库中成功地筛选出识别凝血酶的 Aptamer。修饰了的 Aptamer 的一级结构和二级结构完全不同于相同序列的天然 DNA 或 RNA，因此，胸腺嘧啶 5 位的修饰可能会影响核苷酸的结构特征。胸腺嘧啶 5 位上以较大基团如苯甲基修饰仍然能够被 SELEX 酶所耐受，仍可进行酶放大反应，这扩大了寡核苷酸库化学修饰的可能性。有趣的是，亦可引入氨基侧链产生假肽/核酸库。

某些修饰的寡核苷酸能被光激活产生活性基团，而活性基团能与某些分子共价结合。一

类特殊的 Aptamer(光交联 Aptamer),短暂的紫外光暴露后,Aptamer 能与靶分子发生交联作用。用含有 5－溴尿嘧啶残基和 5－碘尿嘧啶残基的修饰库筛选出了光交联 Aptamer。5′替代胸腺嘧啶介导的光交联发生在核苷碱基 5′替代基团与富电子氨基酸之间。容易发生光交联的氨基酸有色氨酸、酪氨酸、组氨酸、苯丙氨酸、半胱氨酸。因此,要求光交联的氨基酸残基与 Aptamer 的 5′修饰基团在空间几何形状上有较严格的配合。这种光交联的严格性能增加 Aptamer 与蛋白作用的特异性。在光交联过程中利用长波紫外光,如 5－溴尿嘧啶用 308 nm 波长的光,5－碘尿嘧啶用 325 nm 波长的光,可以大大降低蛋白质和 Aptamer 的光降解。虽然可以从 5 位修饰的寡核苷酸库中筛选光交联 Aptamer,但也可以在 SELEX 完成后,进行尿嘧啶的 5 位修饰。

无须光照即可发生交联的 Aptamer 是被偶然发现的,虽然这些 Aptamer 能发生无光交联,但它们仍需要 5－碘尿嘧啶。无光交联 Aptamer 易于在 5－碘尿嘧啶的 C6 位与蛋白质的亲核基团之间形成 Michael 加合物。这种无光交联 Aptamer 拥有光交联 Aptamer 的一切优点,但无须光照。

对诊断分析来说,最重要的是配体与靶分子之间识别的亲和力和特异性。在异种分析中,非特异的分子将被冲洗掉,不会影响 Aptamer 与靶分子的结合反应。在理想状态下,所有的非特异性结合被消除,而特异性结合都不受影响。交联反应的启动在 Aptamer 与靶分子之间形成共价结合,然后可以用极端的冲洗条件去除非特异结合的分子,以大大降低分析的背景。在同种分析中,共价交联能克服可能发生的靶分子分离。

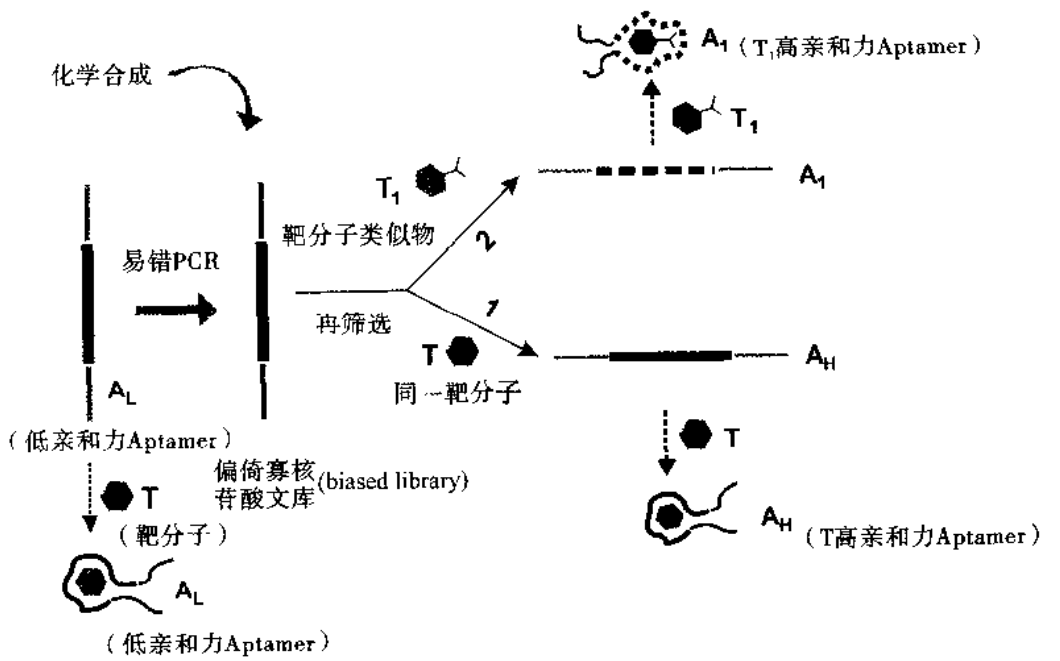


图 9－4　Aptamer 的优化过程

注:建立一个与某一靶分子(T)低亲和力的 Aptamer(A_L)偏倚文库(biased library),可提高 A_L对靶分子的亲和性或改变它对靶分子类似物(T_1)的亲和性和特异性。以 A_L的初始序列为基础,通过控制加入 4 种三磷酸核苷的比例,用组合化学或易错 PCR 控制各位点核酸的比率,从而建立 A_L偏倚文库。另外,还可在 PCR 过程中寻找恰当条件引入随机突变位点。易错 PCR 过程中,以 A_L初始序列为模板,由聚合酶以每位点 0.6%的错误概率引入核苷酸,形成非配对寡核苷酸,组成以 A_L序列为基础的偏倚寡核苷酸文库。偏倚库可再用于筛选同一靶分子的 Aptamer,即 A_L的衍生物 A_H,路线 1 所示,亦可用于筛选靶分子类似物,路线 2 所示。

如上所述，RNA 库经过 2′位的修饰后具有核酸酶抗性，克服了 RNA Aptamer 的一些固有缺陷，可分析含有核酸酶体液中的靶分子。与未修饰的 RNA 相比，未修饰的 DNA 通常具有较好的核酸酶抗性。实际上修饰的和未修饰的 DNA Aptamer 均已用于体液中靶分子的分析，因此未经修饰的 DNA Aptamer 亦可用于诊断。Aptamer 能在体液中作用较短时间而不被降解，如果必要，也可以在 Aptamer 的末端加上一些小分子的修饰，如氨基接头、磷酸基团或转化的胸腺嘧啶残基等，这些末端修饰基团不仅可以保护 Aptamer 免遭核酸酶的降解，也可作为一个连接臂，因为在大多数分析中需要一个连接臂将 Aptamer 与报道分子相连，或通过连接臂固定于载体上。

在单克隆抗体技术中，抗体的筛选其实就是杂交瘤细胞的筛选。如果手中的抗体不是最好的，就必须重新筛选，或者重新制备单克隆抗体，因为对已有的抗体进行改进是十分困难的。而对 Aptamer 的改进则容易得多，可在已有的 Aptamer 基础上筛选出亲和力与特异性更好的 Aptamer，即在已有的 Aptamer 序列基础上产生一个偏倚库，并对偏倚库进行 2 次筛选。偏倚库的分子复杂性较初始序列更为复杂，它包括了初始 Aptamer 的各种变异，但比初次 SELEX 过程所用的随机库更具目标性。如上所述，随机库的复杂度没有理论计算的那么丰富，而且在初始 SELEX 过程中这些最佳 Aptamer 可能会错失，偏倚库提供了另一种筛选最佳 Aptamer 的途径。从概念上讲，这一方法是属于“先导优化”策略（图 9－4），是药物化学常用的策略，即根据先导化合物筛选潜在的化学药物。偏倚库技术已成功用于优化 Aptamer 和核酶。这一方法还可以改变已有 Aptamer 的特异性，如在一个能识别 L－丝氨酸的 Aptamer 序列的基础上，利用偏倚库技术可重新筛选出可特异识别 L－精氨酸的 Aptamer，这种 2 次筛选出来的能特异结合 L－精氨酸的新的 Aptamer 已丧失了结合 L－丝氨酸的能力。这一有趣的例子说明了可以根据需要进一步提高或改变 Aptamer 特性的可行性。

三、影响 SELEX 技术的因素

为获得一次成功的筛选，有几个因素是非常重要的：文库的设计、靶蛋白的浓度、循环次数及分离方法。

（一）文库设计

1．文库的种类：第一个 SELEX 实验是用随机 RNA 文库完成的。Gold（1995）在其综述中统计了 50 个 SELEX 实验，发现其中有 36 个使用的是 RNA 文库，5 个是 DNA 文库，9 个是修饰核苷酸文库（表 9－1）。目前还没有足够证据表明，筛选某一靶分子到底该用 RNA 文库还是 DNA 文库。只有极少数实验同时使用 RNA 文库和 DNA 文库筛选相同靶分子，而且结果表明，筛选到的寡核苷酸配基的亲和力和特异性相似。从寡核苷酸分子结构来看，单链 RNA 分子中除 G－C，A－U 配对外，还有 G－U 变偶碱基对的存在，因此更易形成多样空间结构，可能更有利于与靶分子结合，这也许是大多数 SELEX 文献都采用 RNA 文库的原因。

值得注意的是，不同类型的寡核苷酸文库筛选同一靶子所得到的配基序列是完全不一样的。例如，用单链 DNA 文库筛选的凝血酶的特异结合配基序列属于 G－四联体结构。如果用 5′－（1－戊炔基）－2′－脱氧尿嘧啶替代该 DNA 文库中的胸腺嘧啶，则筛选出的特异配基序列与非修饰 DNA 文库完全不一样。若改用 RNA 文库，则筛选出的特异核酸配基为茎环结构，只与凝血酶肝素结合位点相互作用。实验还表明，用这 3 种不同类型的寡核苷酸文库筛选出的凝血酶特异配基皆有抗凝血作用。

不同的文库有不同的制备方法。RNA 文库可以由含有 T7 启动子的双链 DNA 模板转录产

生，而单链 DNA 文库根据需要不同则有不同的制备方法：

(1)链 DNA 模板变性：将双链 DNA 模板库热变性以后快速冷却，大多数序列将不会找到各自的互补序列，取而代之的是形成一些分子内二级结构的序列。然而，用这种方法产生的单链 DNA 文库存在一些问题。大量的、不同的 DNA 序列中的一些固定序列往往可以复性连接成链状或高分子凝聚物，这些副产品在今后的 PCR 扩增中将会得到扩增而影响筛选。另外这些单链的 DNA 随着后来的不断富集，更易找到其互补序列而重新复性形成双链 DNA。鉴于以上原因，单链 DNA 文库最好来源于双链 DNA 互补链中的一条。

表 9－1　SELEX 筛选的高亲和力序列特性

靶分子	文库种类	随机序列长度(碱基数)	循环数	基序	解离常数 (10^{-9} mol/L)	丰度
T4 DNA 聚合酶	RNA	8	4	发卡*	5	4
噬菌体 R17 衣壳蛋白	RNA	32	11	发卡	5	9
大肠杆菌 rho 因子	RNA	30	8	发卡	1	10
大肠杆菌 S1 蛋白	RNA	30	13	假结	4	12
大肠杆菌核糖体 30S 颗粒＋S1	RNA	30	13	假结	5	12
大肠杆菌核糖体 30S 颗粒－S1	RNA	30	12	SD 序列	7	11
Qβ 复制酶	RNA	30	11	假结	5	12
大肠杆菌 met J 蛋白	RNA	40	15	不明	1	不明
HIV－1 Rev 蛋白	RNA	32	10	凸环	1	11
	修饰的 RNA	37	13	发卡/凸环	1	
HIV－1 tat 蛋白	RNA	40	10	发卡/凸环	5	11
HIV－1 int 蛋白	RNA	30	18	复合型	20	12
HIV－1 逆转录酶	RNA	32	9	假结	5	10
	ssDNA	35	15	茎环/凸环	4	12
MMLV 反转录酶	RNA	36	17	发卡	10	11
AMV 反转录酶	RNA	36	12	发卡	2	11
FIV 反转录酶	RNA	40	18	发卡	2	13
RNA 结合蛋白 U1A	RNA	10	3	发卡*	5	4
	RNA	13	3	发卡*	500	4
	RNA	25	3	无	500	3
RNA 结合蛋白 Het－N1	RNA	25	3	富－U 结构	不明	3
凝血酶	RNA	30	12	发卡	9	11
	修饰的 RNA	30	15	发卡	100	13
	ssDNA	60	5	G－四聚体	25	4
	修饰的 ssDNA	20	6	发卡	400	6

续　表

靶分子	文库种类	随机序列长度(碱基数)	循环数	基序	解离常数(10^{-9} mol/L)	丰度
弹性蛋白	修饰的 RNA	40	15	G－四聚体	15	14
	ssDNA	40	17	G－四聚体	10	14
分泌型磷脂酶 A2	RNA	30	12	复合型	1	11
	修饰的 RNA	50	11	发卡	1	11
神经生长因子	RNA	30	10	假结	100	12
碱性成纤维细胞生长因子	RNA	30	13	发卡	0.20	11
	修饰的 RNA	30	11	不明	0.30	12
血管内皮细胞生长因子	RNA	30	13	发卡/凸环	0.20	12
	修饰的 RNA	30	11	发卡	0.30	9
噬菌体 T7 gp10 蛋白抗体	RNA	10	3	发卡*	不明	6
系统性红斑狼疮抗体	RNA	30	13	发卡	3	13
	ssDNA	40	8	不明	1	10
胰岛素受体单抗	RNA	40	11	发卡/凸环	2	11
IgE 类抗体	修饰的 RNA	40	9	G－四聚体	35	9
人绒毛膜促性腺激素	修饰的 RNA	40	16	发卡	60	14
P 物质	RNA	60	24	发卡/凸环	190	14
维生素 B12	RNA	72	8	假结	88	14
茶碱	RNA	40	8	茎环/凸环	110	14
瓜氨酸	RNA	74	7	茎环/凸环	650	12
ATP	RNA	120	8	茎环/凸环	700	10
活性绿染料	RNA/DNA	100,120	6	发卡	60 000	8,11
D－色氨酸	RNA	120	7	不明	18 000	不明
L－精氨酸	RNA	25	7	茎环/凸环	300 000	9
L－缬氨酸	RNA	25	9	茎环/凸环	12 000 000	9

* 号表示使用的文库为限制性文库，譬如随机序列两端的固定序列部分对称，使之只能形成发卡结构；丰度表示该序列在起始随机文库中所占几率，如 10 表示几率为 10^{-10}(摘自 Gold,1995a)。

(2)生物素末端标记法：此方法是 Bock(1992)等筛选凝血酶的 DNA 配体时所采用的。具体方法是将模板的一条引物 5′端标记上生物素，经 PCR 扩增产生的混合物通过一个亲和素标记的琼脂糖柱，未标记上生物素的 DNA 被洗脱分离。标记上生物素的 DNA 留在琼脂糖柱上，再经碱变性(0.15 mol/L NaOH)，DNA 双链打开，非生物素标记的一条单链 DNA 就被洗脱下来，由此可制备单链 DNA 文库。

(3)不对称 PCR：此技术是 Ellington 和 Szostak(1992)筛选染料分子 DNA 配基时所采用的。不对称 PCR 有多种方法，多是使其中一条引物量远大于另一条引物，经过最初的指数性扩增，由过量引物产生的单链 DNA 线性堆积。这种方法的优点是得到的单链 DNA 产物不需要进一

步纯化。

(4)逆转录:单链随机的 DNA 文库可由该库的随机 RNA 序列逆转录得到。这种方法在制备大规模的起始库时是非常有用的,但在每一轮筛选中却非常麻烦。

2. 两端固定序列。一个典型的 RNA SELEX 所用单链 DNA 模板(40N cDNA)及其引物序列如下:

5′GCCTCTTGTGAGCCTCCTGTCGAA 3′

5′GCCTGTTGTGAGCCTCCTGTCGAA (40N) TTGAGCGTTTATTCTTGTCTCCC 3′ 40N cDNA

3′AACTCGCAAATAAGAACGAGGGATATCACTCAGCATAAT 5′

从以上序列可以看出,寡核苷酸文库的模板中间含有一定长度的随机序列,5′及 3′两端为固定序列,这些固定序列长度一般为 15 ~ 30 个核苷酸不等。SELEX 所用的引物与文库模板两端固定序列相互补。因此,文库固定序列的设计一定要考虑引物的设计对 PCR 的影响,主要是避免 PCR 扩增中引物二聚体和二级结构的形成及引物的自身折叠。如果一个筛选需要 10 轮 SELEX 过程,而每轮 PCR 要进行 30 个循环,那么每个目的序列最终要经过 300 次循环扩增。因此,合理的引物设计,对减少 PCR 扩增中引起的错配或非特异条带十分重要。在遵循上述设计原则的前提下,固定序列中可以适当引进较多的 GC 含量,以提高后面的 PCR 反应过程中退火的温度,增加反应的特异性。

另外,固定序列的设计对文库的稳定性也有重要影响。由于 RNA 酶的广泛存在,并且其作用并不需要特异性核苷酸序列,因此 RNA 文库稳定存在是实验的前提。文库的稳定性主要取决于序列两端的固定序列。目前文献中采用的 RNA 文库 5′端和 3′端的固定序列往往有一段互补区,如:在筛选与 RNA 酶 P 结合的 RNA 配基时,文库两端的固定序列 5′端为:ACGGAUUCGCUCAA,3′端为:UUCAGAAUUCGCACCA(带下划线的序列为互补序列),这使 RNA 文库在空间上能形成一段双链区,从而使其稳定性增加,这也是利用了某种 RNA 文库能够在体内稳定存在的性质设计的。上述筛选 RNase P 所用的 RNA 文库的固定序列就是来自 pre - $tRNA^{phe}$ 的三叶草的叶柄部分。目前为了增加随机 RNA 文库的稳定性,一些实验中对固定序列中的某些碱基进行了修饰,如用 2′ - NH_2 或 2′ - F 取代嘧啶碱基中的 2′ - OH,用 5 - (1 - 戊炔基) - 2′ - 脱氧尿嘧啶替代 ssDNA 文库中的胸腺嘧啶核苷酸等。

除了以上要考虑的因素,固定序列的设计还要考虑 T7 RNA 多聚酶转录的起始。为了便于体外转录,一般在文库模板的固定序列中加入 T7 RNA 多聚酶启动子。另外,为了便于今后的克隆与测序,在文库模板两端固定序列中还要引入限制性内切酶酶切位点。当然,这些酶切位点也可以在最后一轮筛选中通过 PCR 引物加上。

3. 中间随机序列。在 SELEX 技术中,中间随机序列的设计也是影响筛选成败的关键。其中随机序列的长度和随机度是最重要的两个因素。

一个未修饰的寡核苷酸文库,只有 A、C、G、T 或 A、C、G、U 4 种碱基组成。因此,如果文库中的随机序列由 n 个碱基组成,那么该文库就会包含 4^n 个不同序列,也即文库的库容量为 4^n。譬如一个随机序列为 30 个碱基的寡核苷酸文库,其库容量大约为 10^{18}。文库的容量对能否筛到高亲和力的配基至关重要,一般达到 $10^{14} \sim 10^{15}$(随机寡核苷酸数为 22 ~ 25),即可满足实际需要。文库的设计不能只考虑增加库容量。从理论上讲,随机序列越长,库容量越大,而且更易形成大的更为精确的结构。但是,由于合成的文库浓度是一定的,随机序列太长反面使文库中每一种寡核苷酸的丰度降低,从而使特定配基序列的筛选概率降低。当然,随机序列的长度也不能太短,否则导致整个库容量变小,同时也不易形成与靶分子结合所必需的特定空间结

构。目前SELEX技术常用的文库中随机寡核苷酸的长度一般是30个碱基左右，此长度的寡核苷酸已足以形成发卡、口袋、假结、G－四联体等已知的单链寡核苷酸的基本空间结构。体外筛选所用的文库随机序列的长度应当反映被筛选配基的长度。在野生型配基核苷酸序列已知的情况下，随机序列长度的选择就会变得相对容易。譬如，野生型HIV－1的RBE(rev－binding element)长度大约30个碱基，因此理论上与Rev结合的寡核苷酸配基的长度应在30个碱基左右。Tuerk等(1992)用了随机序列为32个碱基的一个文库筛选能与Rev结合的更高亲和力的配基，取得成功。如果单从寻找寡核苷酸药物的角度考虑，则应尽量避免长片段。

随机度是指寡核苷酸序列随机化的程度。随机寡核苷酸可以是完全随机组合、部分随机组合或者被固定的碱基序列分隔成几个小的区段(block)，极少数是将几个随机组合的区段连接成一个非常大的随机区域(多于200碱基)。随机度与SELEX实验目的有关。一般来讲，随机序列应与天然序列保持一定的相似性，这样才能更好地确定到底哪种碱基和结构与天然核酸功能有关。通常对天然小分子核酸(少于20碱基)可以完全随机化，对较大的核酸可以部分随机化，并在中间引入一些非野生型的碱基以区分不同的功能区。

(二)筛选条件

不同的靶分子，不同的文库(如DNA或RNA文库)，或不同的实验目的(如要筛选出单一最佳配基或它的一组相关序列)所需要的筛选条件都不尽相同。对SELEX技术来讲，面对如此广泛的靶分子范围，很难设计一个快速、通用的筛选程序，一般只能给出一些总的指导原则，其中靶分子浓度、寡核苷酸配基的分离方法及循环次数是最重要的几个方面。

1. 靶分子浓度：靶分子的浓度在一定程度上取决于与之结合的核酸配基。由于高亲和力配基的含量在文库中相当低，因此最初几个循环应使用较高浓度的靶分子，以保证捕捉到尽可能多的配基。后面的循环，由于高亲和力配基经过PCR扩增富集，其浓度已足以与高含量的低亲和力配基竞争，此时应降低靶蛋白的浓度，以加强筛选的严谨性。如果在整个SELEX筛选过程中一直使用高浓度的靶分子，将会减少寡核苷酸之间的相互竞争，导致每一轮富集效率降低，并大大增加筛选出特异配基所需的循环次数。

Irvine等通过实验给出下列公式，能估算出接近最佳的靶分子浓度：

$$T_{no} = [(K_d) + (C_{oligo})](0.7)(BG/CP)^{2/3}$$

其中T_{no}为接近最佳的靶分子浓度，K_d为文库总寡核苷酸与靶分子的大致解离常数，C_{oligo}代表文库总寡核苷酸的浓度，BG为非特异结合RNA所占的比率，CP为与靶分子特异结合的寡核苷酸所占的比率。如果文库总寡核苷酸的浓度远大于K_d，则K_d可以忽略不计。

另一个粗略估计靶分子浓度的方法是，以能结合文库总寡核苷酸5%～10%的靶分子水平为靶分子使用浓度。计算机模拟显示，这种情况下将10^{15}库容量中1拷贝的寡核苷酸分子富集到占文库一半以上的含量，所需要的筛选不超过7个循环。

在实际操作过程中，SELEX实验起始寡核苷酸与靶分子的比率一般为10～100，在继后的筛选过程中逐渐增加该比率以提高竞争力。

2. 配基的分离方法：分离与靶分子结合及未结合的寡核苷酸也是SELEX实验中非常关键的步骤，其中主要涉及分离介质、缓冲体系、非特异竞争剂及反应平衡时间。

对不同的靶分子所采用的分离介质也是不一样的。筛选一些小分子，如小肽，通常用固定化的配基制成亲合层析柱，当然也可以将小肽偶联到一些分离介质上，如活化磁珠等。对一些大的蛋白分子，可将蛋白质分子固定在特殊修饰的纤维素滤膜上。如果能获得靶分子的抗体，

将蛋白质－配基复合物以免疫沉淀的形式与游离配基分开也是切实可行的。需要注意的是，很难用滤膜的方法分离与蛋白质结合的 DNA 配基，因为脱氧核糖的强疏水性会带来过高的背景吸附。一般改用其他介质，如用多孔聚丙烯酶联板或电泳分离。

通过控制洗脱缓冲液的条件也能增加分离的严谨性。高浓度的一价阳离子常可抑制文库与靶分子的非特异结合；高浓度的镁离子（或其他二价阳离子）能促进核酸空间结构的形成，当然这会使靶分子的特异和非特异结合有所增强。

加入一些非特异的靶分子结合物可以与核酸配基竞争结合的靶分子，有利于从随机文库中筛选出高亲和力的配基。这些非特异竞争剂可以是 tRNA，也可以是细胞总 RNA。如果要筛选一些特殊表位或功能域的核酸配基，还可加入较特异的竞争剂。如在筛选 HIV－1 的 Rev 结合配基时，可加入 HIV 的 RBE，这样筛选出来的配基能完全避免与野生型 RBE 的结合。对多表位的蛋白质靶分子，竞争抑制某一表位有可能增加配基与其他表位的结合，但在大多数情况下配基只与单一的最主要的表位结合，而且筛选的特异性也可通过控制文库与竞争剂的比例达到。

另外，文库与靶分子的结合需要一定时间才能使反应达到平衡，因此反应平衡时间也是影响分离效果的一个因素。那么，达到平衡需要多少时间？一个较简单的判断规则是：竞争解离常数在 10^{-9} mol/L 左右时，一般反应需要数分钟；竞争解离常数在 10^{-12} mol/L 左右时，则需要数小时才能达到平衡。

最后，在考虑影响配基分离的因素时，还要特别注意文库与分离介质结合引起的背景过高，这是进行 SELEX 筛选时经常遇到的问题。实验发现，文库中经常含有可与分离介质（如硝酸纤维素膜、琼脂糖等）相结合的寡核苷酸配基，而且丰度可能比与靶分子特异结合的核酸配基还高。过高的背景往往增加筛选中的循环次数，甚至导致筛选失败。因此在开始第一轮筛选前，需进行一步反筛操作，即将未结合靶分子的寡核苷酸先过一遍分离介质，以除去与分离介质相结合的部分，降低筛选背景。

3. 循环次数：循环次数依实验而定，通常比设想的要多一些，它与高亲和力配基在文库中的含量有一定关系。如前所述，正确选择靶分子的浓度，降低分离过程中的背景可有效减少一个成功筛选所需的循环次数。对高亲和力配基来讲，每一个循环能得到 10 倍的富集。譬如，在 T4DNA 聚合酶 SELEX 实验中，从 65 000 分子中筛选出 2 个特异序列需要 4 个循环；在 R17 外壳蛋白 SELEX 实验中，从库容 10^9 的文库中筛选出 1 个特异序列大约需要 11 个循环。一般经过很少几轮筛选很难找到丰度很低的高亲和力配基。在 Gold(1995a)的综述中，50 个成功的 SELEX 实验所用的循环多在 10 个左右，个别也有只用 3 个循环或多于 20 个循环的，最终所需的循环次数要由实验中设立的结合力检测实验确定。

第三节　核酸配基结合的特异性评价

一、配基结合特异性

SELEX 的靶范围很广，几乎任何一种可以想像得到的蛋白质、肽类物质、小分子有机物，甚至复杂的混合物，都可筛到与之相匹配的寡核苷酸配基。目前已经报道筛选出特异配基的靶分子多达几十种，而且特异性极强。利用 SELEX 技术筛选出的配基特异性可以用下面的例子说明。茶碱是缓解急性或慢性哮喘的药物，但具有严重的毒副作用，需定期检查血清中的含量以预防毒副作用的发生。茶碱与其他黄嘌呤类似物咖啡因、可可碱的结构非常相似，而后两者

常存在于血清中。常规的茶碱单抗检测有交叉反应，而通过 SELEX 方法筛选到的寡核苷酸配基仅特异地结合茶碱，与其他两种物质无反应，其中与茶碱的亲和力比与咖啡因的亲和力高 10 000 倍。核苷酸、氨基酸和寡核苷酸配基能将它们与突变体、镜像体区分开来。Famulok 和 Szostak(1992)利用随机 RNA 文库筛选出的 D－色氨酸配基与 D－色氨酸的结合率达 80%以上，而与 L－色氨酸的结合率低于 4%。有人还用 SELEX 技术筛选出能分辨辅酶 I(NAD)和还原型辅酶 I(NADH)的寡核苷酸配基。

核酸结合蛋白的寡核苷酸配基与天然结合序列的结构相似，但亲和力更强。一种逆转录病毒 Rev 蛋白的寡核苷酸配基与其他逆转录病毒的 Rev 蛋白不发生相互结合。普通蛋白质比核酸结合蛋白更容易筛选到高亲和力、高特异性的拮抗性核酸配基，K_d一般在 $10^{-9} \sim 10^{-12}$ mol/L 水平。例如碱性成纤维生长因子(bFGF)和血管内皮细胞生长因子(VEGF)是两种与肿瘤血管生成有关的生长因子，均为肿瘤细胞的生长提供营养，它们的特异性拮抗剂可遏制恶性肿瘤的生长。利用 SELEX 筛选得到的这两种生长因子的寡核苷酸配基均能拮抗其与细胞表面受体的结合，而且特异性很强，亲和力比与其他的肝磷脂结合蛋白或变性分子的结合力高 100～10 000 倍。

寡核苷酸配基为何有如此高的亲和力，并能够如此精确识别靶分子呢？它与靶分子之间相互作用的基础又是什么？将获得的寡核苷酸序列输入计算机，结构分析的结果显示，SELEX 筛选出的与靶分子相互作用的寡核苷酸配基大小一般在 10～15 碱基，甚至更小，这些寡核苷酸常存在于一些分子间结构中，如发卡、G－四聚体的凸环、假结(pseudoknot)或口袋中。这些结构中的许多碱基对并不遵循 Watson－Crick 配对原则。核苷酸残基之间具有强烈的相互作用，即使是非常短的一段寡核苷酸，也能形成稳定的结构。一个足够容量的寡核苷酸库包容了所有可能的立体结构，在与靶分子结合时，几乎不需要丢失"熵"就能存在一个适配的结构，因此才有可能从中筛到各种靶分子的高亲和力配基。配基与靶分子的结合部位的平均大小为 300～400 Å，大体与抗体 Fab 片断识别抗原位点大小相当，在此范围内结合只需很少的自由能即可达到解离常数为 10^{-12} mol/L 水平。单个氨基酸之间很少有相互作用。短肽在溶液中很难形成固定的结构，只有比较大的蛋白质才能形成稳定的立体结构，但是其侧链之间也缺少很强的相互作用。因此，与其他组合化学库例如随机肽库、抗体库相比，寡核苷酸库的优越性显而易见，这也回答了为什么从随机肽库、环肽库，或者将随机序列引入一个预定的抗体结构中所筛到的配基的亲和力和特异性都不如核酸配基的问题。

二、配基序列及空间结构分析

分析配基序列及其结构对理解配基与相应靶分子的结合特异性很有帮助。

在用随机 RNA 文库筛选 HIV－1 逆转录酶特异配基的实验中，文库所用的随机序列长度为 32 个碱基。经过 9 轮 SELEX 筛选，筛选出的核酸分子与 HIV－1 逆转录酶功能位点结合的 K_d为 5×10^{-9} mol/L，并能抑制 cDNA 的合成，而对其他病毒逆转录酶无影响，说明筛选出的配基为 HIV－1 逆转录酶特异的。而且分离了 104 个克隆，测序后发现其中包含 20 种特异序列。通过排列、对比分析，证明序列中存在两个明确的一致序列基序(motif)：CGGGA(U/A)A 和 AAUCCG(U/A) A，并发现前一序列模式在 104 个克隆的分子中高度保守。这些序列并不相互重叠，中间被 4～11 个碱基隔开，多为 6 个碱基。前面 18 个序列明显可归为一类分子，也即同一家族(family)，后面两个序列不包含上述两个功能基序，它们是与硝酸纤维素膜特异结合的核酸分子。计算机分析其二级结构，显示其为一假结(psueodoknot)结构。上述例子说明，配基

序列及空间结构分析可获得许多重要信息。

SELEX 技术的建立在于随机文库中含有大量的具有不同分子构象的寡核苷酸分子。由于分子结构决定其功能，因此利用 SELEX 筛选出的一组针对靶分子某一特异功能的配基应具有共同的三维结构特征。而具有共同三维结构基序的核酸分子通常也具有相似的二级结构基序。这些二级和三维结构基序在核酸分子序列上表现出一定的模式或一级结构上的相似性。根据这些不同的序列模式，可将筛选出的核酸分子归为不同的家族。

当然，从文库中富集的功能分子往往不是单一的一个家族，有可能是多家族分子的复合物，这是由于文库核酸分子与靶分子不同功能基团相互结合造成的。

将具有共同三维结构的分子序列加以比较，将得到许多与配基功能有关的信息，如保守核苷酸的鉴定、特殊位点突变的可能性等。同一家族中功能重要的区域一般高度保守，而只对核酸拓扑结构如凸环形成有作用的区域变异性较大。另外，配基的分子刚性(rigidity)与其结合的特异性与亲和力也有密切关系。配基分子刚性越强，其与相应靶分子结合的特异性和亲和力越高。这种现象在多肽与受体的结合中也存在。

第四节　SELEX 技术的改进与发展

自 SELEX 技术问世以来，该技术本身的发展非常迅速，许多改良的 SELEX 技术相继建立。

一、导向 SELEX

导向 SELEX(blended SELEX)是将一个已知的可以与靶蛋白上的特定位点结合的小分子掺入寡核苷酸库或者将一个适当的小分子释放到靶蛋白上。这样通过 SELEX 筛选出的配基只与该蛋白上的特异表位结合，增加了药物作用的特异性，避免了它与细胞内一些未知成分的结合所带来的副作用，这一点在药物筛选中十分重要。例如利用带有葡萄糖分子的寡核苷酸库可筛选出己糖激酶的作用位点。Bless 等(1997)从一个掺有弹性酶小分子不可逆抑制剂(缬氨酸二苯脂磷酸)的随机 ssDNA 文库中筛出具有高度底物特异性的弹性酶的 ssDNA 配基，并首次在动物的肺部炎症模型中证明该配基能抑制肺损伤和嗜中性粒细胞的流入。另外，若在合成寡核苷酸库时掺入 2′－NH_2UTP、2′－NH_2CTP，可有效地抵抗人体内的核酸内切酶，增强药物或检测试剂的稳定性。

二、复合靶分子 SELEX

经典 SELEX 筛选的靶分子多为单一蛋白或小分子物质，通过这些靶分子筛选出的寡核苷酸序列一般仅有很少几个家族。这些家族的配基常相互竞争靶分子的同一位点或交叉位点，而这些位点通常又是靶分子上占优势的位点。有趣的是，实验发现当数种靶分子混合在一起时，这些优势位点仍然优先与寡核苷酸文库结合。例如，运用 SELEX 技术以 30 S 核糖体颗粒作为靶子筛选出的 S1 蛋白的配基与用单一纯 S1 蛋白作为靶子筛选出的配基，在序列及亲和力上都极为相似，似乎表明文库与各个不同位点的结合并不互相干扰。Morris 等(1998)在此基础上将 SELEX 技术用于人红细胞膜总蛋白配基的筛选，建立并完善了复合靶分子 SELEX(complex targets SELEX)技术方法。该方法与经典 SELEX 不同之处主要有两点：一是筛选的靶分子并非一单纯蛋白或小分子，而是人红细胞膜的总蛋白，为多种靶分子的混合物，实际上，它是利用同一个寡核苷酸库同时进行多个筛选，将混合物中不同靶子的特异配基同时筛选出来而彼此互不干扰；二是必须将筛选出的多种靶分子特异配基进一步分离，所用方法是将筛选出的特

异配基复合物在交联剂的作用下与原复合靶分子(如上述人红细胞膜总蛋白)交联,同时标记上同位素,经 SDS/PAGE 电泳分离因靶分子大小不同显示的核酸－蛋白条带,PCR 扩增后再重复上述操作,最终得到每一靶分子各自的特异配基。

复合靶分子 SELEX 是近两年发展起来的非常有实际应用价值的一项技术。它所针对的靶子不是纯的蛋白质或小分子,而是多种物质的混合体,血清、尿样、动物血管、肿瘤细胞、红细胞、白细胞、细胞浆或细胞核抽提物等都可以作为它的靶子,应用范围大大扩展。另外,设计一种适当反筛选实验(如我们正在建立的消减杂交 SELEX 技术),可以获得识别两种混合体之间的差异寡核苷酸配基。如果选用的混合体是一个肿瘤细胞和一个正常细胞或两个不同发育阶段的细胞,则可通过 SELEX 技术发现新的肿瘤标志、发育标志及不同的胞内信号分子,这可能成为新的诊断或治疗靶子。用随机肽库研究复杂的生物系统有一定的局限性,因肽类配基的亲和力较弱,不可能完全检测到混合物中所存在的靶蛋白,而寡核苷酸配基则能将混合物中含量极微的靶蛋白寻找到,并研究其功能。

三、基因组 SELEX

蛋白质与 DNA 及 RNA 的相互作用是基因表达调控的核心问题,有些蛋白质在基因表达调控中甚至与两种或更多的核酸发生相互作用。目前已知,这种蛋白质与核酸的调节网络十分复杂,许多问题还不十分清楚。对大多数蛋白质来说,利用经典 SELEX 能筛选出与之结合的 RNA 配基,配基亲和力可达 10^{-9} mol/L 水平,足以引起体内生物效应。但是这些蛋白质多数并不属于 RNA 或 DNA 结合蛋白,因此需要建立一种可以广泛筛选 DNA 和 RNA 天然配基的技术。

基因组 SELEX(genomic SELEX)就是能满足这一要求的新技术。基因组 SELEX 是在经典 SELEX 技术的基础上衍生来的。它将感兴趣的生物体的基因组制备成一个寡核苷酸文库,从中筛选蛋白质、抗生素、辅因子、多糖等生物活性分子的天然识别序列,为解析细胞内基因调控、代谢调控的问题提供大量的数据材料。基因组 SELEX 技术成败的关键是起始文库的质量。理想的文库应包含所感兴趣基因组的所有序列。基因组文库的构建可用多种方法,如限制性酶切和连接法、机械片段化和平末端连接法及随机引物 PCR 扩增法等。Singer 等应用随机引物 PCR 的方法将 *E. coli*、酵母及人的基因组 DNA 构建成了可用 T7 RNA 聚合酶扩增的基因组文库。

第五节　适配分子的特性

谈到适配分子就不能不谈到抗体,抗体在基于分子识别的许多领域都作出了巨大贡献。从 20 世纪 50 年代开始抗体被用于检测分析,到 20 世纪 70 年代由于多克隆抗血清的发展,抗体分析技术被广泛应用于各个领域,由于需要大量的抗体,多克隆抗体制备技术难以满足需要,单克隆技术的发展使大规模制备独特型抗体成为可能。由于抗体具有巨大的实用价值,单克隆技术被广泛用于世界的每一个角落。单克隆抗体技术具有很多优点,如能从培养的细胞中制备大量的抗体;抗体能被容易地优化和提纯,目前单克隆抗体已被广泛用于各种非同位素分析。今天已经能在连续培养的过程中选择出产生单克隆的细胞克隆。理论上讲,产生抗体的细胞能无限的产生抗体。另一方面,所用抗原不必纯化。单克隆技术的发展给免疫学带来了巨大革命,并大大扩展了抗体的应用范围,但抗体在某些方面仍有不足之处。

1. 抗体的第一步需要用抗原免疫动物。某些分子,如毒素由于动物不能耐受而无法制备

抗体，此外一些固有的免疫原性较差的分子也很难制备抗体。

2. 杂交瘤细胞只限于大鼠和小鼠，这限制了单克隆抗体在治疗中的应用，非人源性抗体主要用于体外诊断，抗异源性抗抗体（人体中识别非人源性抗体的抗体）能使检测出现假阳性，此外风湿因子和自身抗体亦影响免疫分析。

3. 制备单克隆抗体是一项实验技巧很强的技术，从大量的克隆中筛选稀有抗体将是一件十分繁杂而昂贵的事情。

4. 筛选出来的克隆细胞在保存过程中有可能因意外而丢失或发生细胞死亡。一般在制备单克隆抗体过程中，将杂交瘤细胞种植于动物腹膜腔，然后从腹水中纯化抗体。但有些杂交瘤细胞很难在体内生长，从而影响了抗体的制备。

5. 有些抗体批与批之间存在差异，因此在免疫分析中需要排除这些影响。

6. 抗体是在体内产生的，而免疫分析是在体外完成的，体外抗体识别抗原是处于非生理状态。

7. 抗体与抗原相互作用的靶分子不能根据需要进行调整。

8. 抗体对温度十分敏感，易发生不可逆变性，抗体的储存期限也有限。

虽然杂交瘤抗体有以上缺陷，然而人们也找到了各种克服的办法，包括人源性抗体的制备技术、噬菌体肽库技术、核糖体技术、抗体工程、体外免疫等。

另一方面，人们可以考虑另一类完全不同的分子—Aptamer，它可以弥补抗体的不足。Aptamer 具有以下优点：

1. Aptamer 的确定是通过体外过程完成的，它不依赖于细胞或动物，也不在体内，其特性可根据需要改变。

2. 确定 Aptamer 的选择条件可以控制，以获得具有体外诊断所需特性的 Aptamer，例如，可以制备出在非生理溶液中和非生理温度下结合靶分子的 Aptamer。同样 Aptamer 与靶分子结合的动力学参数也可以改变。

3. 由于不需要免疫动物和细胞，毒素和没有免疫源性的分子也可产生高亲和力的 Aptamer。

4. Aptamer 能用化学合成的方法制备，具有很高的精确性和重复性。在变性条件下能纯化得到很高的纯度。在制备 Aptamer 的过程中几乎没有批与批之间的差异。

5. 标记分子如荧光素和生物素能标记到 Aptamer 的精确位置，在合成过程中亦可以衍生上去。

6. Aptamer 能可逆变性。变性后的 Aptamer 能在数分钟之内复原，并能长期储存和在常温下运输。

第六节　SELEX 技术和适配分子在各方面的应用

自 SELEX 技术问世以来发展非常迅速，由于其筛选的靶范围广、亲和力高、特异性强，因此在基础研究、临床诊断和药物筛选等领域中展示了其广阔的应用前景。

一、SELEX 技术在各方面的应用

（一）SELEX 技术在基础医学研究中的应用

如前所述，基因组 SELEX 的建立是 SELEX 技术在基础医学研究中应用的典范，利用它可

以筛选出基因复制、转录和翻译过程中起关键作用的许多蛋白调节因子或其特异拮抗剂，借以研究其调控机制。它为研究核酸－蛋白质的相互作用提供了一种非常有效的研究手段。例如，细胞内存在一些特异的活性 DNA 片段，乳铁传递蛋白（lactoferrin）只有与 DNA 片段发生相互作用后才表现出传递活性。而在体外，自然杀伤细胞（NK cell）中乳铁蛋白的活性能被一些 RNA 或 DNA 抑制。因此用 SELEX 技术筛选出的乳铁传递蛋白的特异核酸配基非常有用。

除此之外，由于 SELEX 技术本身是利用了随机 RNA 文库筛选靶分子寡核苷酸配基，因此近年来成为研究核酶最方便、最直接的手段之一。利用 SELEX 技术不仅能筛选出核酶的特异寡核苷酸配基及其剪切位点，而且还可以在随机 RNA 文库中发现一些新的核酶序列。譬如，酵母 $tRNA^{phe}$ 与 Pb^{2+} 共同孵育时能进行自我剪切，Pb^{2+} 的剪切速率与 tRNA 分子的适应折叠密切相关。Pan 等（1992，1995）将 Pb^{2+} 剪切反应引入 SELEX 中，成功地分离到 tRNA 样折叠的新 RNA 分子，而且还从中找到在 Pb^{2+} 存在下能进行自我剪切的 RNA，也即新的核酶。根据 SELEX 原理，还可将已知核酶中的保守序列随机化，构建成随机 RNA 文库，从中筛选出比野生型活性更强的核酶，并对其生物活性及功能进行研究。Pruden（1994）筛选出的 RNA 具有催化桥联式双苯分子异构体互换的功能，使该酶的概念有了新的扩展。

SELEX 技术还可用于新蛋白质的细胞内定位及分离。由于寡核苷酸配基与蛋白质结合的亲和力在 $5\times10^{-11}\sim5\times10^{-8}$ mol/L 之间，高于抗原与抗体的结合，因此将生物素、荧光素、活性基团偶联到经 SELEX 筛选出的靶蛋白特异寡核苷酸配基上，可以对新蛋白质进行细胞内定位，也可以将这些特异配基作为层析介质，制成亲和层析柱分离纯化该蛋白质。近几年的数据累计表明，人们在寡核苷酸配基和抗体方面所作的努力，是朝着一个相同的方向迈进的。鉴于 SELEX 技术的优越性，可以预言，在不久的将来，特异寡核苷酸配基很可能成为主动免疫中抗体的替代物，实验室里的单克隆抗体制备技术将会被一种费用不高、较为方便的 SELEX 技术所代替。

近年来，随着核磁共振技术及计算机技术的飞速发展，通过解析经 SELEX 技术筛选出的特异配基同源家族一致序列的结构，为揭示生物大分子间的相互作用规律提供了大量的实验数据。Robison（1998）将 55 种 *E. coli* DNA 结合蛋白的结合位点制成矩阵，与 *E. coli* 基因组 DNA 共同孵育，得到的天然结合位点与经 SELEX 技术筛选出的 DNA 结合位点进行比较，建立了关于这些蛋白与 DNA 相互作用的数据库。最近，SELEX 技术常被用于研究 RNA 与蛋白质之间的相互作用，如鉴别筛选能与锌指蛋白特异结合的 5 S 核糖体 RNA 最小功能域及与一些冷休克蛋白（cold shock protein）相互作用的 RNA 分子的特异序列等。

（二）SELEX 技术在药物筛选中的作用

筛选治疗性药物是 SELEX 技术发展早期的主要目的。例如，阻止肿瘤血管生成是防止肿瘤生长的有效手段之一。肿瘤细胞分泌大量的血管生成刺激因子如血管内皮细胞生长因子（VEGF）、成纤维细胞生长因子（FGF）等，肿瘤血管内皮细胞也过量表达一些关键受体。利用 SELEX 筛选这些生长因子或受体的特异拮抗剂可以阻止肿瘤血管的生成，起到抗肿瘤作用。目前，利用 SELEX 筛选出的 VEGF 拮抗剂已经进入临床试验阶段。此外，经 SELEX 筛选出的凝血酶的寡核苷酸配基能抑制凝血酶的作用，防止血栓形成；全身性红斑狼疮病人的体内产生了大量抗自身抗原的抗体，属自身免疫性疾病的一种，用 SELEX 筛选出的抗体的 RNA 和 ssDNA 配基，能拮抗这些自身抗体的作用；CD4 分子的 RNA 配基与鼠淋巴细胞表面 CD4 分子的亲和力高于 CD4 单抗与 CD4 分子的结合，并且能以 CD4 特异的方式完全阻断同种异体混合淋巴细胞反应（MLR）；活化血小板表面的细胞黏附分子 P－选择素（P－selectin）介导血小板与中性

粒细胞的结合，在早期炎症反应中起到重要作用。P－选择素的寡核苷酸配基的 K_d 为 10^{-11} ~ 10^{-10} mol/L，是其糖蛋白类特异配基的 10^5 ~ 10^6 倍，是 E－选择素和 L－选择素的 10^4 ~ 10^6 倍，成为炎症治疗的候选药物之一。另外 L－选择素的 ssDNA 特异配基也能阻断体内 L－选择素的功能，具有潜在的临床应用价值；Goodman(1999)利用 SELEX 技术筛选出能与宿主细胞 DNA 整合的细菌蛋白 IHF 的特异配基，为细胞感染性治疗提供了新的思路。Biesecker(1999)用该技术获得人补体 C5 的特异抑制剂，经化学修饰后得到的 2′－F RNA 特异配基的生物活性不变，很有可能用于临床治疗与补体 C5 有关的人类疾病如慢性感染等。另外，其他一些病毒蛋白、细胞因子的寡核苷酸配基也有一定的治疗前景。相信 SELEX 技术在药物筛选中的作用会越来越重要。

如果 SELEX 筛选出来的寡核苷酸配基特异性及稳定性能满足临床要求，那么这些配基可以直接作为药物用于疾病的治疗。但是利用 SELEX 筛选出来的寡核苷酸配基尽管与靶分子的亲和力很高，其稳定性往往达不到要求，因此从临床应用角度考虑，还需对寡核苷酸配基进行各种修饰，以进一步提高其特异性和稳定性。修饰方式有两种，一种是在筛选前，将修饰性核苷三磷酸掺入寡核苷酸库。常用的修饰位置是磷酸基团、嘧啶环 5 位、嘌呤环 8 位以及五碳糖的 2′位。另外一种是在筛选出的寡核苷酸配基上进行上述修饰。而在合成寡核苷酸库时掺入 2′－NH_2UTP、2′－NH_2CTP，可有效抵抗人体内的核酸内切酶，增强药物或检测试剂的稳定性。随着计算机药物设计的发展，将 SELEX 筛选出来的特异寡核苷酸配基进行序列测定，通过分析其空间结构，从小分子化合物数据库中寻找类似结构的化合物作为先导物，进一步发展成稳定而特异的药物用于临床。

(三)SELEX 技术在临床诊断方面的应用

由于寡核苷酸配基的高特异性和高亲和力，将其作为诊断试剂比本质为蛋白质的抗体类诊断试剂具有更高的灵敏度和准确性，而且这种特异性和亲和力不受组织或样品中的非靶蛋白干扰。如系统性红斑狼疮是一种自身免疫性疾病，病人的体内产生大量抗自身抗原的抗体，并随病情的变化而消长。其中一种抗体被用作 SELEX 的靶子，经 SELEX 筛选出特异抗体的 RNA 和 ssDNA 配基，不仅可用于系统性红斑狼疮的诊断，还可用于病情监测。相信随着 SELEX 技术的不断完善，许多很难找到特异诊断试剂的疾病可以通过 SELEX 技术完成。前面讲过的将 SELEX 技术用于哮喘病人血清中茶碱的检查而排除具有类似结构的咖啡因、可可碱的干扰，也是 SELEX 技术用于临床检验的典型范例。

随机寡核苷酸文库几乎可以作用于所有的物质，对于一些未知的靶分子，如流行病中的一种新的病毒蛋白或一种新的中毒介质，可以不考虑其结构和特性，直接用 SELEX 技术筛选出它的特异配基作为诊断试剂，或经过改造用于预防与治疗，这一点在今后军事医学研究中将会有十分重要的价值。

另外，以同位素标记特异寡核苷酸配基作为显示剂，还可用于体内造影，以诊断局部病灶型病变。

二、Aptamer 在各方面的应用

(一)Aptamer 在双位点分析中的应用

双位点分析又叫夹心分析，是一种最常用的诊断分析方法。在分析中被测物被夹在两个配体之间，一个配体用于捕获，另一个用于检测，原理如图 9－5 所示。Aptamer 被成功用于血

管内皮源性生长因子(VEGF)夹心分析，一 DNA Aptamer 被用作检测配体，然后利用碱性磷酸酶连接的抗荧光素抗体 Fab 片段(商业化产品)可检测出 5′端荧光素标记的 RNA Aptamer。在这一分析中也可以用酶连接的 Aptamer 代替酶连接的抗体 Fab 片段。这一分析可检测出血清中的 VEGF。另一个成功应用的例子是利用 Aptamer 和抗体同时结合细胞表面的 CD4 分子，且由于两个配体的结合部位在 CD4 杆状分子的不同部位，故二者互不影响。在这两个分析中 Aptamer 都没有被固定在固相载体上。Aptamer 被固定在微球上，可在夹心分析中捕获靶蛋白分子。由此可见，Aptamer 在夹心分析中即可用作捕获配体，也可用作检测配体。

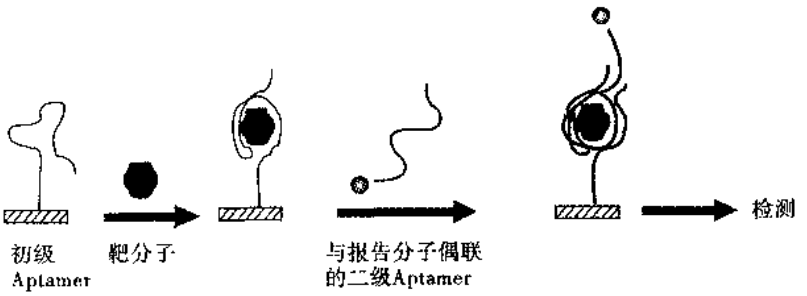


图 9－5　双 Aptamer 夹心分析法

注：在该方法中，有两个 Aptamer。初级 Aptamer 固定在固相支持物上，与靶分子结合形成靶分子－Aptamer 复合物。二级 Aptamer 一端标有恰当的报道分子(如荧光素)，特异地识别复合物。通过检测报道分子变化(如荧光)情况来检测靶分子的水平。

到目前为止，基于 Aptamer 的夹心分析都是用一个抗体作为第二配体，这是由于难以找到两个结合位点不同，且结合位点没有重叠的 Aptamer，其原因是由 SELEX 过程本身所决定的。标准的 SELEX 过程常是为了确定亲和力最高的 Aptamer，亲和力是整个过程所追求的惟一目标。若亲和力的强化达到了极限，最后产生的是亲和力极高的、结合在最佳部位的 Aptamer。如果限制 SELEX 的过程，也许可以找到两个不同结合部位但亲和力有所下降的 Aptamer。此外也可以改变策略来寻找结合部位没有重叠的 Aptamer。已经筛选出来的 Aptamer 可以用做后续 SELEX，以找到不同部位的 Aptamer，也可以将靶分子以不同的形式呈现，进行两次不同的 SELEX 过程，以找出针对同一靶分子不同表位的不同 Aptamer。但不用上述技术也可找出两个针对同一分子不同结合部位的 Aptamer。Lochrie 筛选出了既可结合 HIV 病毒复合蛋白中壳包核酸蛋白部位又可结合基质蛋白部位的 Aptamer。两类 Aptamer 互不竞争结合部位。Tasset 等分离出来的结合人凝血酶的 DNA Aptamer 与 Bock 分离出来的凝血酶 Aptamer 具有不同的结合部位，但两个凝血酶 Aptamer 的 DNA 序列有惊人的相似，两个 Aptamer 都能适应凝血酶分子内部的 G－四重结构，但一个 Aptamer 结合在凝血酶的纤维蛋白原识别部位，另一个则结合在凝血酶的肝素识别部位。如上所述，产生两个 Aptamer 的 SELEX 过程不尽相同，其中在一个 SELEX 过程中，凝血酶溶于溶液中，以单链 DNA 作为库来筛选，而另一个则将凝血酶固定在固相载体上，固定的靶分子也许将另一个结合位点遮掩了。Lin 等用两个不同结合位点的凝血酶 Aptamer 设计的一个纯 Aptamer 夹心分析实验取得了较好的结果。

Aptamer 怎样与靶分子作用是一个令人感兴趣的研究领域，利用核磁共振光谱技术阐明了部分 Aptamer 与小分子的作用机制。这些实验揭示：①Aptamer 能折叠成独一无二的能容纳靶分子的口袋形结构；②Aptamer 上的一些功能基团能与邻近靶分子上的对应基团相互作用；③

Aptamer 区分分子结构的能力达到了原子水平。一些小分子与其 Aptamer 复合物核磁共振实验表明，Aptamer 能将小分子完全包埋在其形成的口袋中，因此小的靶分子就没有机会再跟其他的第二配体结合，从而限制了小分子的第二个 Aptamer 的筛选。

另一种实验对于利用 Aptamer 分析小分子很有价值，即筛选出针对小分子 - Aptamer 复合物的 Aptamer。由于核酸可相互作用，因此可相信能分离出针对靶分子 - Aptamer 复合物的 Aptamer。这种 Aptamer 对基于 Aptamer 的小分子的夹心分析很有帮助，故可以完全不需要抗体进行小分子及大分子蛋白的分析。在抗体分析中这种利用抗体来识别靶分子与抗体所形成的复合物的方法叫做"抗体免疫复合物免疫分析"。该方法被成功地用于血清样本中地高辛的分析，在分析中二抗被用于特异识别地高辛与第一抗体所形成的复合物。虽然这一设想得到了普遍认可，但在实际分析中存在困难，主要是由于很难分离出针对小分子 - 第一抗体复合物的第二抗体，筛选针对靶分子 - Aptamer 复合物的第二 Aptamer 将大大推动这一分析方法的应用。

Aptamer 的高度特异性在诊断分析中必须仔细考虑。Aptamer 能分辨靶分子的细微结构的变化，如靶分子上的甲基、羟基，以及尿素和胍基，对靶分子的手性亦有很高的分辨率。Aptamer 的特异性水平与单克隆抗体相似。单克隆抗体已成功用于区分靶分子的细微差别，如利用单克隆免疫分析检测糖基化蛋白以及来源于不同组织的同工酶。这种高度特异性的分析也可以用 Aptamer 完成。但高度特异性的单克隆抗体在分析一些具有分子多态性的靶分子如垂体糖蛋白激素时遇到了麻烦，该麻烦在 Aptamer 的分析中也可能遇到。但这一问题可用"鸡尾酒"Aptamer 予以克服。鸡尾酒 Aptamer 能特异地识别靶分子的变异，它们可以很快制备，特别是利用全自动 SELEX 更容易。这对于具有分子变异性的靶分子的分析更有实用价值。

酶的催化能力常用来增强诊断分析的敏感性。在 ELISA 分析中，第二抗体多与酶如碱性磷酸酶或辣根过氧化物酶连接起来，由酶催化产生信号。在这里虽然不讨论 Aptamer 的催化特性，但利用这一技术有人已发现具有催化活性的寡核苷酸。在将来可能会采用具有催化活性的寡核苷酸来产生信号(图 9 - 6)，从而将蛋白质挤出舞台。最近，Wilson 等分离出了一种结合磺酸化罗丹明荧光球的 Aptamer，作者发现这种 Aptamer 能使荧光球中的物质发生氧化而产生荧光，虽然其催化能力较弱，但可进一步优化 Aptamer 以达到最佳的催化效力，利于诊断分析。

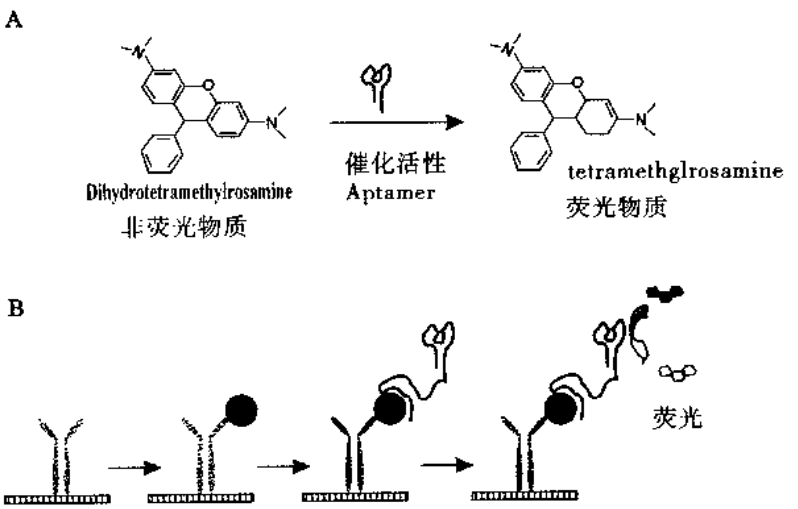


图 9 - 6　催化寡核苷酸在体外诊断中的应用

注：Aptamer 可催化荧光底物发生变化，发出荧光。(A)非荧光物质 dihydrotetramethylrosamine 在催化 DNA Aptamer 作用下，产生 tetramethylrosamine，后者含有 π－电子对，能够发出荧光。(B)催化 Aptamer 在体外诊断中应用的延伸。该方法类似 ELISA 法。在双位点结合分析中，催化 Aptamer 为二级配体。一级配体与催化 Aptamer 结合，后者催化荧光发生物产生荧光。

(二)Aptamer 在流式细胞分析中的应用

流式细胞技术是一种强有力的分析工具，能分析单个细胞或细微颗粒的多个参数。现在该技术已广泛用于基础研究和临床诊断。诊断分析建立在能结合配体或待测分子的微球颗粒上，由于其根据微球颜色和大小进行区分，故流式细胞技术已成为一项颇具吸引力的分析平台。装备有先进信号处理软件的流式细胞仪能在几秒钟内检测几千个颗粒中含有的微量物质。若在带有荧光染料的微球上连接一定的配体，如抗体或核酸，则能在流式细胞仪上进行多参数分析。

目前，已利用流式细胞仪和特异的 Aptamer 来分析微球或细胞表面的某种特异蛋白。例如，将人类嗜中性弹性蛋白酶(HNE)DNA Aptamer 与荧光素连接起来，可取得比单克隆抗体更强的信号。利用 2－氟嘧啶 Aptamer 可检测人类细胞上的 CD4 表达情况。也可将抗体与 Aptamer 结合起来，同时分析细胞表面多个靶蛋白分子。

利用荧光素标记的剪切 DNA Aptamer 成功地检测了细胞表面的炎症素。针对炎症素的 Aptamer 能粘附在白细胞和中性粒细胞上，利用 Aptamer 的流式细胞分析结果与单克隆抗体的分析结果十分相似。加入未标记的 Aptamer 能完全抑制中性粒细胞粘附的标有荧光素的 Aptamer，但仅能部分阻断抗炎症素单克隆抗体与中性粒细胞结合。一种能结合 P 炎症素的 Aptamer 也分离了出来，是一种 RNA Aptamer，能结合在表达 P 炎症素的血小板上。

IgG 抗体是双价的，而典型的 Aptamer 是单价的。为研究双价 Aptamer 的行为，合成了双价 Aptamer 并利用流式细胞仪研究其特性：二聚性。Aptamer 可通过固相化学合成对称的二价 Aptamer，二价 Aptamer 的亲和力是单价 Aptamer 的 10 倍以上。二聚化以后的 L 炎症素 Aptamer 的亲和力是单价 Aptamer 的 20 倍。双价 Aptamer 的离解率仅为单价 Aptamer 的 1/10。双价 Aptamer 在细胞表面的半衰期长于双价抗体一倍以上。由此可见 Aptamer 的性能可不断提升。

Aptamer 的流式分析揭示了一些有趣的现象。连接有荧光球，或荧光素，甚至结构巨大的蛋白如分子量达 240 KD 的藻红素的 Aptamer 仍保留很好的结合特性，这说明 Aptamer 能容纳各种各样的报道分子，以适应诊断分析的需要。可选择最佳的报道分子以得到最大的信号强度。Aptamer 既可完全替代抗体也可与抗体联合使用，产生双价 Aptamer，甚至多价 Aptamer，这种方法十分方便，且可大大提高 Aptamer 的特性。由于抗体存在 Fc 片段，而有些细胞上又有 Fc 受体，因此影响了抗体在某些细胞中的应用，而 Aptamer 在这方面有更多的优势。

(三)Aptamer 在传感器中的应用

有时需要在很短的时间内完成分析，甚至做到实时分析。基于分子识别的传感器能满足这一需要，理想的亲和传感器应满足以下 3 个条件：①在传感分析中无须另加试剂；②在很短的时间内检测出所需浓度范围的待测物质；③能多次重复使用。

抗体免疫传感器是免疫分析的另一种形式，在免疫传感器中抗体被固定在传感器的表面与待测物质结合。免疫传感器的一个重要缺陷是表面抗体的再生能力较差，需要较温和的条件来保持抗体的活性。虽然免疫传感器可多次使用，但抗体的失活不可避免。而基于 Aptamer 的传感器则有很多优点，首先是 Aptamer 的再生能力。由于 Aptamer 是核酸，它能反复变性和复性，可用许多方法对 Aptamer 进行再生，如加热、改变盐的浓度、改变 pH 或加入螯合剂等，极

端的条件也不会破坏 Aptamer(但极端的 pH 应避免,以免 Aptamer 被水解)。再者,Aptamer 容易被修饰,便于固定在传感器的敏感膜上。将核酸固定在不同表面的方法已经十分成熟,操作简便且重复性很好。在 Aptamer 的化学合成中可以加上一个连接臂,可精确地将 Aptamer 按所需的密度固定在表面。Aptamer 的第三个优点是能标记各类报道分子,因此可以设计多种检测方法,将荧光球附着在 Aptamer 上,当 Aptamer 与靶分子结合时,荧光球的荧光特性将会改变。下面将具体讨论这一特性的应用。SELEX 技术的另一个重要优点是可以通过改变 SELEX 过程中的条件而改变 Aptamer 的特性,例如,从含有荧光球修饰的随机寡核苷酸库中能筛选出具有特殊荧光特性的 Aptamer,这种 Aptamer 可应用于传感器上。

最近,Aptamer 已被应用于多种亲和传感器。如将生物素化的能识别 L－腺苷的 RNA Aptamer 固定在链菌蛋白素衍生物的光纤表面制成传感器上,固定后的 Aptamer 的结合特性与溶液中的结合特性一致。该传感器具有很高的手性分辨率,它对 L－腺苷的亲和力是 R－腺苷的 1 700 倍。利用识别凝血酶的 Aptamer 构建了检测凝血酶的消失波荧光异向性传感器,Aptamer 的 5′端用荧光素标记,3′端用烷基氨标记以便附着在玻璃表面,该传感器的优点是不需对待测靶分子进行标记,因此可用于体内测量,此传感器具有很高的特异性,它仅对凝血酶有响应,而对待测溶液中的胰肽酶没有响应,而杂乱的 DNA 用荧光素标记后固定在传感器表面对凝血酶也没有响应,说明该传感器是在 Aptamer 与靶分子特异性结合的基础上响应的,它有很好的重复性。将表面的 Aptamer 变性后即可将已结合的凝血酶去除,可使传感器重复使用。虽然目前 Aptamer 在传感器方面的应用仍十分有限,但 Aptamer 在亲和传感器的应用方面有巨大潜力。

另一个尚未探索的 Aptamer 在传感器方面的应用是 Aptamer 结合的电化学检测。由于 Aptamer 是多聚阴离子,当与靶分子结合后,其导电性将发生改变,电荷在核酸双螺旋中的转移规律是一个热门的研究领域,虽然这一研究仍处于起步阶段,但具有巨大的应用潜力。目前电荷转移仅限于纯的 DNA 范围内,但不久将会研究 DNA 与其他分子相互作用时的电子转移规律。Aptamer 的应用也将随着新发现的诞生而进一步扩大。

(四)Aptamer 在荧光偏振分析中的应用

荧光偏振分析常被用来分析小分子物质,如类固醇物质。在竞争分析中标记的类固醇与待测类固醇共同竞争抗体,荧光偏振(FP)的变化直接反映了标记分子与抗体的结合状态,待测的未标记的分子与抗体结合使 FP 信号减弱。竞争分析快速简便,且有较好的重复性。

但竞争分析存在一些缺陷,如灵敏度不够高,且动力学范围较窄,需要精确控制试剂的用量。抗体的固有特性决定了抗体的荧光偏振分析必需采用竞争分析法。结构较大的抗体分子在溶液中的旋光较慢,因此抗体旋光率的变化不能明显反映与小分子的结合情况,因此,荧光标记的抗体与靶分子结合后其 FP 信号变化很小甚至检测不到。此外均匀的荧光标记的抗体制备也比较困难。而 Aptamer 的情况则不一样,基于 Aptamer 的荧光偏振分析(如图 9－7),荧光素可标记在 Aptamer 的尾部,而不必标记靶分子,因面可采用非竞争分析。荧光标记的 Aptamer 在 FP 分析中可有 2 种形式:①由于 Aptamer 的分子较小,只有抗体分子的 1/10,因此,其旋光率变化较快,因而 FP 信号变化较大;②核磁共振试验结果显示 Aptamer 结合小的靶分子后会产生构像变化,也会有旋光率的变化。

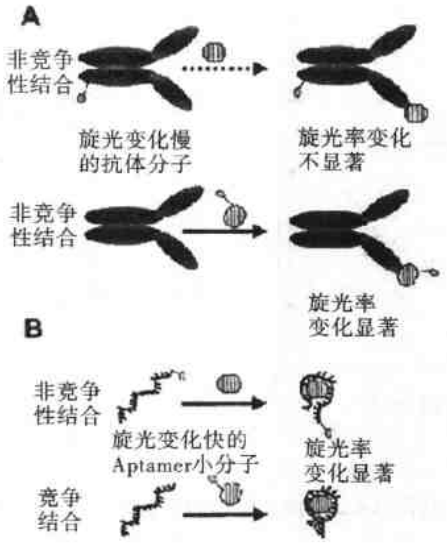


图 9－7　利用抗体(A)和 Aptamer(B)进行荧光偏振分析检测小分子

注:小分子与另一标有荧光素的分子结合,产生旋光率变化,旋光率变化与荧光偏振(FP)的变化直接相关。(A)用大量荧光素标记的抗体进行非竞争性(NC)同质物分析是不可行的。因为荧光素标记的抗体很大,与小靶分子结合后,其旋光率变化不明显,故不易检测其变化。在竞争性(C)分析中,将靶分子进行荧光标记,其与非标记的抗体结合,由于抗体很大,故旋光率变化很显著。

(B)与抗体相比较,Aptamer 两种情况都适用。非竞争性分析中,Aptamer 较小(约为抗体的 1/10),与靶分子相互作用时,可产生显著的荧光偏振变化。

利用荧光标记识别 HNE 的 DNA Aptamer 对 HNE 进行荧光偏振分析取得了较好的结果。如图 9－8 所示,当标记的 Aptamer 与 HNE 结合后,FP 信号有明显的变化,而加入较高浓度的非特异性人血清白蛋白的 FP 信号变化则不明显,说明可利用荧光标记的 Aptamer 进行非竞争性的荧光偏振分析;利用识别氨基糖甙类抗生素的 Aptamer 进行竞争性荧光偏振分析显示了 Aptamer 与氨基糖甙类抗生素的特异结合;利用荧光素标记的凝血酶 Aptamer 构建凝血酶生物传感器,当传感器表面有凝血酶时,由消失波激发的荧光偏振就会发生改变。凝血酶生物传感器在 140 pL 的样本中可检测 0.7～700 amoles 的凝血酶。与凝血酶分析相比,基于 Aptamer 的 HNE 分析是在溶液中进行。当 Aptamer 与靶分子结合时,标记在 Aptamer 分子上或标记在靶分子上的荧光素的荧光特性都会改变,这对 Aptamer 在荧光分析中的应用十分有用。实际上,识别黄素蛋白的 Aptamer 能淬灭靶分子的荧光,因此 Aptamer 能用于本身有荧光特性的靶分子的分析。

(五)Aptamer 与分子灯塔的结合

最近,一种叫分子灯塔的荧光基因探针用于同源核酸序列的检测。分子灯塔是一个发卡状探针,在其一端标有荧光素,而在另一端则标有荧光淬灭基团,如图 9－9 所示。当分子灯塔折叠成发卡状时,荧光基团所发出的荧光将被淬灭基团所淬灭。发卡状灯塔分子的核苷酸序列设计成与靶序列互补,靶分子与灯塔分子杂交后,环状结构打开,淬灭基团被移开,原来没有荧光发射的灯塔分子将发射荧光。利用分子灯塔检测了致病相关的核苷酸序列。到目前为止,灯塔分子主要用于检测核酸。因为尚无与灯塔分子特异结合的非核酸靶分子,如蛋白靶分子。

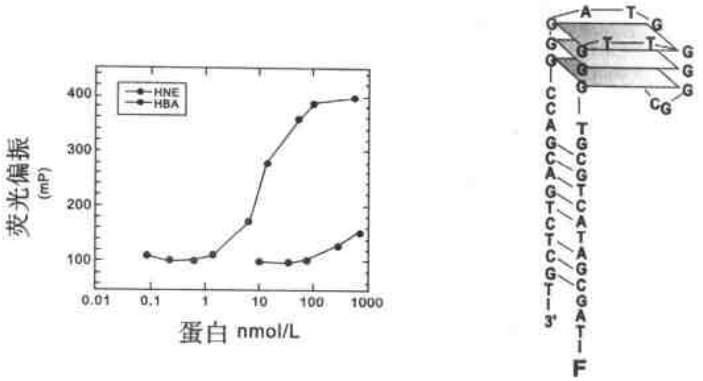


图 9-8　利用 Aptamer 荧光偏振分析检测 HNE

注：右图为荧光素 5′标记的 Aptamer 结构。左图为荧光标记的 Aptamer 与血清球蛋白或 HNE 在特定缓冲体系中孵育，蛋白浓度与荧光极性的变化情况。

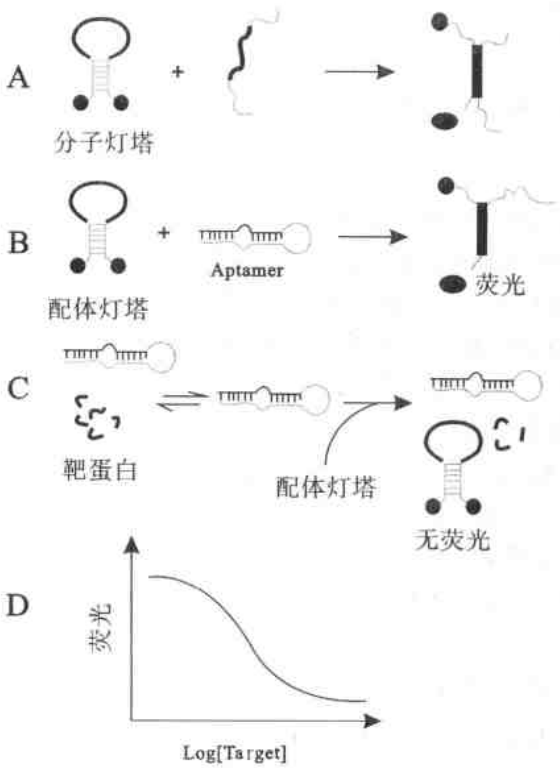


图 9-9　基于 Aptamer 的分子灯标

(A)用于检测特异同源核苷酸序列的分子灯塔。(B)配体灯塔指能够与 Aptamer 相连产生荧光的分子灯塔。图中黑色环为识别 Aptamer 靶位点区的序列。配体灯塔与 Aptamer 结合，产生荧光。(C)与靶蛋白分子结合的 Aptamer 不能再与配体灯塔结合。因此，当靶蛋白过量时，配体灯塔由于不能和 Aptamer 结合，故灯塔不亮。(D)用分子灯塔法分析靶分子浓度：当靶分子浓度极低时，所有 Aptamer 与配体灯塔结合，产生很强荧光；当靶分子浓度极高时，所有 Aptamer 都与靶分子结合，配体灯塔无法与 Aptamer 结合，不产生荧光；当靶分子浓度在二者之间时，荧光强度与靶分子浓度成反比。

Aptamer 既能与核酸分子相互作用，也能与非核酸分子作用，因此可以利用 Aptamer 将核酸灯塔分子与非核酸分子连接起来，建立灯塔分子与非核酸靶分子之间的桥梁。与 Aptamer 相互作用的分子灯塔也叫配体灯塔。配体灯塔的环状序列与 Aptamer 的部分序列互补。Aptamer 的互补序列仅在没有靶分子的状态下才能与灯塔分子的环状序列杂交。当靶分子存在时，配体灯塔不能与 Aptamer 杂交，利用这一原理可设计基于 Aptamer 和配体灯塔的竞争分析，该分析既可分析缓冲液中靶分子，也可分析血浆中的靶分子，具有广阔的应用前景。

同时分析多个物质的分析方法具有很多优点，利用多个分子灯塔，每个灯塔分子连接上一个不同的荧光物质，每个荧光物质发出不同波长的荧光，这样可以分析多个核苷酸靶序列。该方法也可用于配体灯塔，可在同一试管中同时分析多个非核酸靶分子。但不同的是它利用体积较大的荧光小珠，以便进行流式细胞分析。利用流式细胞平台可同时分析多个蛋白质组分，这一技术也叫固相配体灯塔分析。由于荧光物质在固相小珠上，而不在液相中，不需要配体灯塔中的淬灭基团，使整个系统更简便。

在这一分析中配体灯塔取代了与 Aptamer 非特异作用的分子，大大提高了分析的特异性。此外，配体灯塔也可能干扰特异的 Aptamer－靶分子结合，特别是当 Aptamer 与靶分子的作用变弱时显得尤为突出，因此那些与靶分子有较低离解率的靶分子特别适合于配体灯塔分析。光交联 Aptamer 能在靶分子与 Aptamer 分子之间形成共价结合，有利于配体灯塔分析。显然配体灯塔分析所提供的这种独一无二的分析方法是抗体无法替代的。

（六）Aptamer 在毛细管电泳中的应用

毛细管电泳（CE）与普通电泳原理一样，但是在毛细管内进行的。该方法具有很多优点：快速、样本体积小、适合于自动化操作、灵敏度高、可进行多成分或多参数分析、在 CE 中组分的分离很快、并可进行在线检测。传统的电泳需要专门的检测手段，而 CE 将分离和检测结合了起来。现在 CE 也被用于免疫分析。抗原－抗体复合物在电场的作用下，在溶液中进行分离，竞争性的和非竞争性的免疫分析都在 CE 上完成。在竞争分析中未标记的待测靶分子与荧光标记靶分子共同竞争定量的抗体分子，而在非竞争分析中过量的标记的抗体与靶分子结合。如前所述，竞争性分析有许多缺陷，而只要条件许可我们一般采用非竞争性分析。但在实际操作过程中，非竞争性分析在 CE 中存在许多问题，当待测分子很小时，很难将游离的标记抗体与标记抗体－靶分子复合物分离开，很多抗体具有同样的电泳特性，这可能与抗体蛋白的糖基化有关。此外同一抗体在标记后的均一性将有所改变，使同一标记抗体的电泳特性不同。

利用 Aptamer 进行亲和探针 CE 可克服抗体的上述缺陷。利用特异识别人 IgE 的 DNA Aptamer 进行亲和探针 CE 很容易将 IgE 分离并检测出来，且加入其他血液成分对分析没有影响。利用硝酸纤维滤膜技术证明，荧光标记的 Aptamer 的离解常数要高于未标记的 Aptamer 6 倍。虽然标记 Aptamer 的亲和力可因标记荧光球而减弱，但最主要的原因是由于电场的作用使 Aptamer 与靶分子分离。亲和力较差的凝血酶 Aptamer 也成功地应用于亲和探针 CE。

Aptamer 具有一致的电荷/质量比率，因其电泳行为可预测。当 Aptamer 与靶分子结合时会发生结构改变，同时质量亦发生变化，因此其电泳特性也发生了改变。由于 Aptamer 的带电特性，在其与靶分子结合时，电势变化十分敏感。Aptamer 技术的优点是可根据需要筛选相应特性的 Aptamer，例如，选择条件和分离方法可根据需要确定；可给予一定的电场。另一方面，CE

也可作为 SELEX 过程中的分离方法，这样得到的 Aptamer 更适合于亲和 CE。但有趣的是 Aptamer 已被成功地用于 CE，而没有利用 CE 来分离 Aptamer。

（七）Aptamer 作为分子开关

Aptamer 的固有特性是可通过改变环境因素反复进行变性和复性，该特性在蛋白质中是不存在的，除非这种蛋白质来源于耐热微生物。这一特性使 Aptamer 能在一定条件下作为分子开关来使用。原理如图 9－10 所示。Aptamer 与其他核酸序列的区别在于它能折叠成一个特定的口袋结构与靶分子结合，而该结合对环境因素如 pH、螯合离子、脱水等的变化十分敏感，使 Aptamer 与靶分子的结合能对上述因素作出反应，因此可根据需要选择上述因素的适当条件。

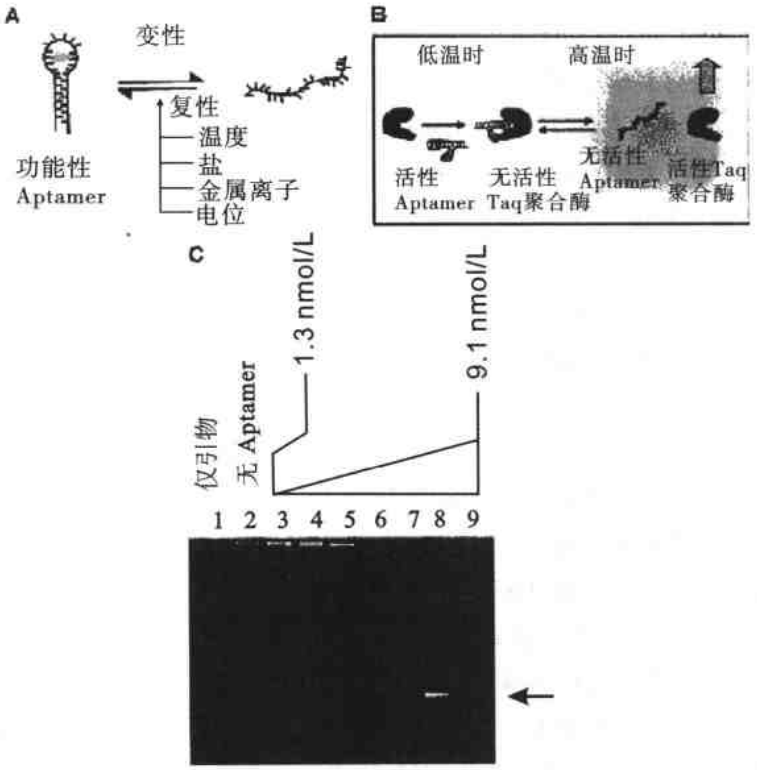


图 9－10　Aptamer 分子开关及其应用

（A）Aptamer 变性、复性可逆转换的示意图及影响因素。Aptamer 的变性复性与外界环境因素有关，这一特性，可使它成为感受外界刺激的传感分子。

（B）温度敏感 Aptamer 原位热启动的 PCR 示意图。通过建立恰当的热启动条件，可建立温度敏感开关，提高了 PCR 诊断的效率、灵敏度和特异性。筛选出低温条件下与 Taq DNA 聚合酶有高亲和力的 Aptamer。将其加入 PCR 反应体系。在低温时，Aptamer 与 Taq 酶结合，影响 Taq 酶活性，阻止低温时引物的非特异性扩增。温度升高时，Aptamer 变性，与 Taq 酶分离，酶活性恢复，进行核酸聚合。

（C）温度敏感 Aptamer 原位热启动的 PCR 实例。扩增人 *k－ras* 基因的一个片段，长 104 bp。孔 1 是未加模板的阴性对照。孔 2 不加 Aptamer。3～9 孔分别加入了 29 ng 从人胎盘中提出的 DNA 模板。低温 Aptamer 浓度由左至右依次增加一倍，孔 3 为 1.3 nmol/L，孔 9 为 9.1 nmol/L。下图为 3% 琼脂糖电泳结果。箭头所示为目的条带。由图可得出，目的条带的特异性扩增需要适当的 Aptamer 浓度。浓度过高，完全抑制了酶的活性；浓度过低，产生非特异性扩增条带，均不能产生理想的扩增条带。

Romig 等利用 Aptamer 进行了亲和色谱分离。其主要是将某金属离子依赖的炎症素 DNA Aptamer 固定在固相载体上，产生了一种 Aptamer 亲和基质，该基质可用于分离炎症素融合蛋白，然后用 EDTA 进行洗脱，这与免疫亲和色谱十分相似。只是将抗体换成了 Aptamer，但这种基质可反复再生，这一点它要优于抗体。

在临床诊断特别在 PCR 诊断中排除污染具有特别重要的意义。在 PCR 中一旦加入试剂，样本管就必须密闭以免污染，如果能对试剂进行"开、关"控制将具有重要意义。在大量的非特异序列中放大拷贝数极少的待测序列常常会导致一些非目标靶序列的放大，非目标靶序列的放大将降低 PCR 的灵敏度，且造成假阳性。这种非特异性放大的原因在于退火时 DNA 聚合酶的非特异性作用，而非特异性引物退火发生在特异性退火最佳温度之下。

热启动 PCR 是一种有效的提高特异性的方法，它是在严格的退火温度达到最佳时才加入扩增的必须试剂。虽然热启动 PCR 能提高灵敏度和特异性，但增加了污染的可能性，且操作复杂，特别是对大量样本更加困难。现发明一种"原位热启动"方法，该方法就是在较低温度下使 Taq 多聚酶丧失活性，在较高温度下被激活。一种能中和多聚酶活性的单克隆抗体被用于原位热启动 PCR，在低温下抗体具有中和能力，而高温下抗体被永久变性，丧失中和能力，而多聚酶被激活。虽然这种单克隆抗体能提高 PCR 的灵敏度和特异性，但对 RNA 靶分子的放大则不适合，这是出于启动多聚酶的温度不适合于 RNA，在高温下 RNA 会降解。RNA 的 PCR 对致病病毒的检测具有重要意义，且具有巨大的商业价值。RNA 的放大需要先利用反转录将 RNA 转变成 cDNA。而反转录的启动需要在较低温度(40～50 ℃)下进行。

Aptamer 能选择性的结合 Taq 酶，且亲和力很高(K_d值为 30～50 pmol/L)，当温度低于 40 ℃时，Aptamer 与 Taq 酶结合时，Aptamer 能抑制酶活性，因此 Aptamer 可用于原位热启动 PCR。Aptamer 对酶活性的抑制是可逆的。而且 Taq 多聚酶的 Aptamer 能结合从 Thermus 属菌分离出来的多聚酶，而不结合其他耐热微生物分离出来的多聚酶。因此，这种 Aptamer 能抑制 Tth DNA 多聚酶，用于 RNA 的原位热启动 PCR。

Taq 多聚酶、Taq 多聚酶的 Stoffel 片段、Tth 多聚酶是 3 种常用的 PCR 酶，Tth 多聚酶用于放大 RNA。Taq 多聚酶是最常用的 PCR 酶，Stoffel 片段对热更加稳定，主要区分错配延伸，Stoffel 片段的这些特性使其在多重 PCR，特异引物 PCR，稀有序列 PCR 中有独特的应用。两种来源于不同序列库剪切的 Taq 酶 Aptamer 对 3 种多聚酶具有不同的特异性，两种 Aptamer 的杂交二聚体对单个多聚酶都有抑制活性，犹如双特异性抗体。两种不同功能的 Aptamer 能结合起来形成双功能 Aptamer，这种双功能 Aptamer 用在原位热启动 PCR 中能控制 3 种酶的活性。Aptamer 的作用能够在 40 ℃以上"关掉"。Aptamer 的多种用途能使其成为温度敏感的分子开关。

筛选库的化学修饰有利于分离出环境敏感的 Aptamer。例如含有离子基团的库有利于筛选对盐敏感的 Aptamer，含有金属螯合基团的库有利于筛选金属离子敏感的 Aptamer。利用不同的库可以筛选出两种不同的针对同一靶分子 HNE 蛋白的 Aptamer。两个 Aptamer 采用相同的筛选方法和条件，一个采用 2′氨基嘧啶 RNA 库，一个采用单链 DNA 库。含 2′氨基嘧啶的 RNA Aptamer 对 pH 十分敏感，特别是在氨基的 pKa(pH 6.2)处更为敏感，而这种亲和力的 pH 依赖性在单链 DNA Aptamer 不明显。含 2′氨基嘧啶的 Aptamer 能对 pH 响应，因此它可作为 pH 敏感的分子开关，如图 9－11。

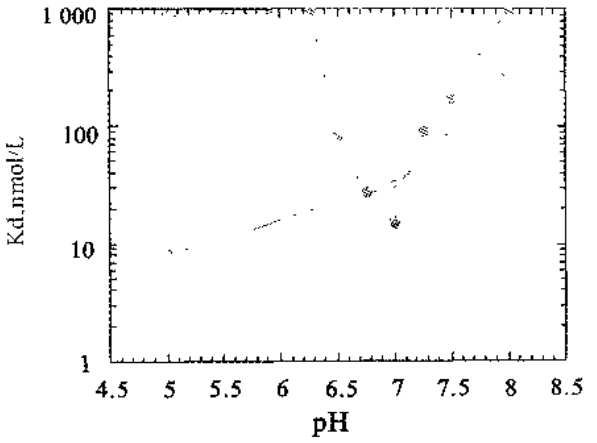


图 9－11　pH 敏感的 Aptamer 及其特性

图为 HNE 与两种 Aptamer 结合率与 pH 值的关系。Aptamer 一种为 DNA(○蓝色),另一种为含有 2′－NH_2的 RNA(●红色)。尽管两种 Aptamer 与 HNE 相结合的条件相同,但含有 2′－NH_2的 RNA Aptamer 比 DNA Aptamer 对 pH 值的变化更敏感。

(八)检测固定在膜上的蛋白质

检测靶蛋白的一种简单方法是利用疏水或亲水作用将蛋白质固定在膜上。蛋白质一旦被固定在固相基质上,就能被标记有报道分子的配体检测出来,当然还需要严格的冲洗以消除非特异性结合带来的背景干扰。在 Northern blotting 和 Southern blotting 标有报道分子的寡核苷酸探针被用于检测固定在膜上的 RNA 和 DNA,在 Western blotting 中被固定在膜上的蛋白质被标有报道分子的抗体检测。由于 Aptamer 能特异、高亲和力结合蛋白质,因此可用标有报道分子的寡核苷酸检测固定在膜上的蛋白质分子,人们称之为"Eastern blotting",完成了 blotting 的全部命名。

下面举例说明 Eastern blotting 的应用。针对人甲状腺刺激素(HTSH)的 Aptamer 被用于检测固定在硝酸纤维膜上的 HTSH。HTSH 属于含糖激素,有许多类似的激素与其结构相似,如:人卵泡刺激素(HFSH),人促黄体激素(HLH),人促绒毛膜性腺激素(HCG),4 种激素的结构十分相似。每种激素都是由 α 和 β 两个亚单位非共价结合而成,4 种激素的 α 亚单位结构都一样,且 β 亚单位也十分相似,进行免疫分析时抗体很难将 4 种激素区分开,而 Aptamer 对 HTSH 有很高的亲和力,且能将其他 3 种结构类似的激素区分开,显示了极高的特异性。这种高特异性的 Aptamer 是从 2′氨基嘧啶寡核苷酸库中分离的。分离该 Aptamer 不需要进行 Counter—SELEX。利用同位素标记的 Aptamer 检测固定在膜上的 HTSH 就具有很好的剂量依赖关系。

利用 SELEX 技术分离出来的以 HCG 为靶分子的 Aptamer 亦具有很高的亲和力和特异性,结合 HCG 的 K_d为 4 nmol/L,能以 400 倍的亲和力区分 HFSH 和 HTSH,但该 Aptamer 能以几乎相同的亲和力结合 HLH,Kd 常数为 HCG 的两倍,这可能是由于二者的结构相似程度太大的缘故。HCG 和 HLH 的 α 亚单位完全相同,而 β 亚单位的 85% 的氨基酸序列相同,HCG 在 C 端多 24 个氨基酸残基,这些差异可用单克隆抗体来区分而没有交叉。在制备单克隆抗体时,用纯化的 β 亚单位作为抗原,可制备出针对 HCGβ 亚单位 C 端的 单克隆抗体;如果用 counter SELEX 则可制备高特异度的 Aptamer,并能区分 HCG 和 HLH;用同位素标记的 Aptamer 检测固定在硝酸纤维膜上的 HCG,二者呈浓度依赖关系,怀孕妇女尿样检测的信号比非怀孕妇女尿样高 100

倍以上。

在 Western blotting 中，经过聚丙烯凝胶电泳(PAGE)后的蛋白质被转移到膜上，然后用标记好的抗体进行检测。Drolet 等用 Aptamer 检测了血管因子 VEGF 的 4 种异构体，异构体是由于 mRNA 的剪切差异所致，用单克隆抗体来捕获 VEGF，然后用 RNA Aptamer 检测，主要检测 VEGF165，但 VEGF121 有一定的干扰。但如果采用 SDS－PAGE 则 Aptamer 仅能结合 VEGF165，而不结合 VEGF121，而抗体对两种异构体都能结合，无法将二者区分开。

(九)Aptamer 阵列研究蛋白组

由于人类基因组计划提前完成，越来越多的基因将被发现，基因组的信息将不断扩大，揭示这些基因所蕴涵的信息资源依赖于预测这些基因的表达产物的功能。显然从 DNA 甚至 RNA 水平预测基因的功能十分困难，这是由于在翻译过程中的剪切将导致蛋白表达的差异，而且翻译之后的修饰对蛋白质的功能亦十分重要，使整个过程变得十分复杂。由于基因型(DNA 或 RNA)与表现型(蛋白质)之间的巨大差异，因此研究蛋白质对于了解基因组在细胞中的功能有意义；对疾病的诊断、疾病进程的检测及疾病对药物的反应更有极其重要的意义。

虽然很容易理解研究蛋白质的重要性，但蛋白质组的研究是一件艰难的任务。它犹如一个巨大的金字塔，其基础在于分析所有蛋白质在组织中的表达及在体液中的分泌状况。微阵列提供了一个同时分辨大量生物分子的手段，如用一个固定在载玻片上的微阵列可同时分析全部酵母的基因组。可利用与基因互补的特异亲和力的寡核苷酸作为微阵列检测 DNA 的基础，但检测蛋白质时需要特异的、高亲和力的蛋白质配体作为蛋白质微阵列检测的基础。虽然已经建立了以抗体为基础的能同时检测大量蛋白质的微阵列技术，但以高亲和力 Aptamer 为基础的微阵列具有许多诱人的优点：

1. Aptamer 能通过全自动平台快速筛选。

2. Aptamer 很容易按一定的密度精确固定在固相载体表面，可用已有的 DNA 微阵列技术制备 Aptamer 微阵列。

3. 可用化学合成的方法同时制备多个 Aptamer。

4. Aptamer 微阵列经久耐用，保存期长。

5. 含有 5′位盐替代尿嘧啶的 Aptamer 能与结合蛋白发生不可逆结合而锁住蛋白。

6. 在 Aptamer 与蛋白质之间形成不可逆交联具有高度的特异性，而且从另一个方面增加了特异亲和力。这种由交联 Aptamer 提供的双特异性使检测时无须使用第二配体，因此可以设想，被光交联 Aptamer 微阵列捕获的蛋白质可用蛋白特异的试剂检测而无需与核酸反应。

核酸性质的 Aptamer 被证明是发展传感器和传感器阵列的十分有效的工具。主要是因为：

1. Aptamer 可形成复杂的三维结构，能以高亲和力和高特异性结合一系列的靶分子。

2. 选择 Aptamer 的方法是比较简单和高效的，可同时对多个靶分子进行高效选择。

3. Aptamer 具有很多与蛋白质分子同样的功能。包括与配体结合后而特异地改变它们的三维结构，这一性质使它很容易转化为传感器的敏感分子。

荧光技术为检测 Aptamer 与靶分子的结合提供了极为敏感的方法。在这些技术中，Aptamer 覆盖在玻片或光纤上，玻片(或者光纤)在内反射的条件下发光。尽管这些光全部由内反射产生，一些能量仍被转移到玻片附近的介质上，形成一个成指数衰减的电磁场。使用该技术可以开发一种方法来识别单一的 Aptamer 分子和靶分子的结合。这两种技术提供了：①一种特异结合一系列靶分子的普遍方法；②一种检测这种结合的极端敏感的方法。可把这些方法结

合起来应用在传感器阵列中进行蛋白质的体外检测。

在不久的将来，Aptamer 微阵列将在蛋白组研究中发挥重要作用，它不仅有利于分析疾病过程中的蛋白质表达情况，而且提供一些新的有效的治疗靶分子。

总之，在过去 30 年中，抗体作为一种识别分子在诊断分析中发挥了重要作用，现代的诊断平台都是根据抗体的特性发展起来的。具有高亲和力和高选择性的 Aptamer 的发现将给目前的诊断分析带来巨大冲击。Aptamer 在不同的分析诊断实验中都取得了令人鼓舞的结果。利用 Aptamer 进行诊断分析能克服许多抗体分析所固有的缺陷。但目前 Aptamer 的研究仍处于青春发育期之前，出现成熟的 Aptamer 产品仍需要一段时间，但对其研究的加速期已经到来。随着自动 SELEX 过程的出现，Aptamer 的鉴定效率将大大提高。值得指出的是 Aptamer 几乎可以满足任何分子识别的需要，它可能被率先使用在那些抗体无法满足要求的领域。今后也可能出现抗体和 Aptamer 联合使用的诊断分析。抗体不能共价捕获靶分子，而 Aptamer 则可轻而易举地共价捕获靶分子。此外 Aptamer 微阵列将在蛋白质组计划中发挥重要作用。

SELEX 技术作为“有形”的物质存在，冲破了传统意义上关于核酸只是作为信息储存和转运载体的理解，让人们认识到寡核苷酸功能的多样性。有人曾提出在前生物时代，世界是一个“RNA 世界”，RNA 既贮存信息，也行使一切生物学功能。后来，随着翻译过程的出现，蛋白质产生了。蛋白质的侧链之间没有强烈的相互作用，蛋白质和蛋白质之间互不“理睬”，在细胞浆 100 mg/mL 的高浓度下，仍能独立地存在，正常地行使功能。寡核苷酸则不同，在 10 μg/mL 的低浓度下，可以自由地折叠成一定形状，而当浓度达到 100 mg/mL 时，它已不可能单独存在，而是与其他分子相互吸引，形成一个胶状物。这种性质上的差异，使得蛋白质在进化过程中占据统治地位，成为生物有机体的重要功能行使者。然而寡核苷酸的特性却适应了迅速发展的组合化学技术的需求。首先，寡核苷酸库可为任何靶分子提供足量的、包纳了所有可能空间结构的底物；其次，寡核苷酸配基与靶子之间的结合高度特异且亲和力很高；另外，SELEX 筛选过程简便、费时短，可较快得到结果；SELEX 筛选出的有治疗前景的寡核苷酸可直接进行动物实验，省去了繁琐的后加工处理；最后，根据 SELEX 筛选出的寡核苷酸配基的结构，还可以利用计算机模拟出新的稳定的药物用于临床。所有这些都将使 SELEX 技术成为一种很有发展前途的组合化学技术。

（徐顺清　王小利）

第十章

基因打靶技术

内容提要

简要介绍了基因打靶技术的概念、发展历史及其在现代生物医学研究中的应用前景。以 *Smad5* 基因剔除小鼠为例，介绍了研制基因剔除小鼠的具体步骤。最后部分是基因打靶研究中常用的实验操作方法。

基因打靶技术是一种定向改变生物活体遗传信息的实验手段，它的产生和发展建立在胚胎干(ES)细胞技术和同源重组技术成就的基础之上，并促进了相关技术的进一步发展。基因打靶技术将广泛应用于基因功能研究、人类疾病动物模型的研制以及经济动物遗传物质的改良等方面。与日俱增的突变小鼠品系及其研究结果将提供越来越多的基因在脊椎动物发育和各种生理病理过程中功能的立体图卷，最终将促进人类对生命现象本质的理解。

第一节　基因打靶技术的概念及发展历史

一、基因打靶技术的概念

基因打靶技术是一种定向改变生物活体遗传信息的实验手段。通过对生物活体遗传信息的定向修饰包括基因灭活、点突变引入、缺失突变、外源基因定位引入、染色体组大片段删除等，并使修饰后的遗传信息在生物活体内遗传，表达突变的性状，从而可以研究基因功能等生命科学的重大问题，以及提供相关的疾病治疗、新药筛选评价模型等。基因打靶技术的发展已使得对特定细胞、组织或者动物个体的遗传物质进行修饰成为可能。它的产生和发展建立在胚胎干(ES)细胞技术和同源重组技术成就的基础之上，并促进了相关技术的进一步发展。它的原理和操作流程如图 10－1 所示。

首先获得 ES 细胞系，利用同源重组技术获得带有研究者预先设计突变的中靶 ES 细胞。通过显微注射或者胚胎融合的方法将经过遗传修饰的 ES 细胞引入受体胚胎内。经过遗传修饰的 ES 细胞仍然保持分化的全能性，可以发育为嵌合体动物的生殖细胞，使得经过修饰的遗传信息经生殖系遗传。获得的带有特定修饰的突变动物提供研究者一个特殊的研究体系，使他们可以在生物活体中研究特定基因的功能。目前，在 ES 细胞进行同源重组已经成为一种对小鼠染色体组上任意位点进行遗传修饰的常规技术。通过基因打靶获得的突变小鼠已经超过

千种(相关数据库参见文献),并正以每年数百种的速度增加。通过对这些突变小鼠的表型分析,许多与人类疾病相关的新基因的功能已得到阐明,并直接导致了现代生物学研究各个领域中许多突破性的进展。

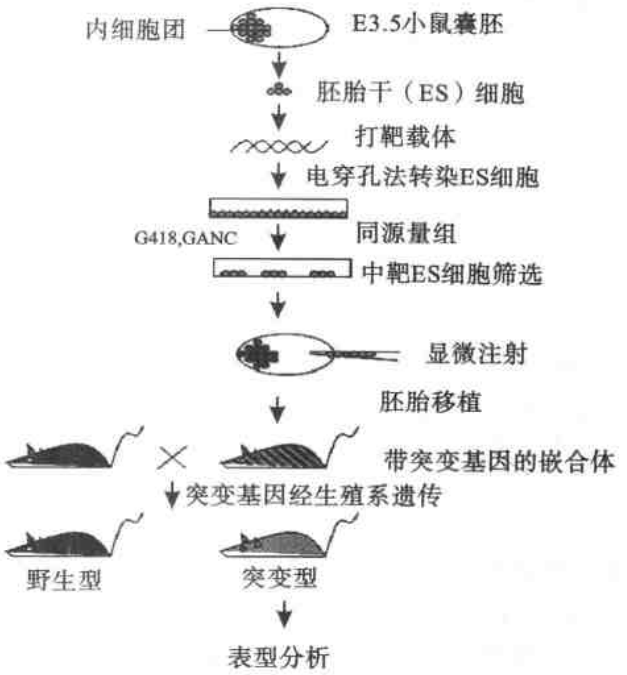


图 10-1 基因打靶的主要技术步骤

二、胚胎干细胞技术的发展

干细胞是指一类具有无限分裂能力的细胞,它们能通过不对称分裂复制自己的同时产生不同类型的细胞。ES 细胞是从着床前的哺乳类胚胎中分离的全能干细胞,具有在体外不分化的增殖能力,经过长期体外培养后仍然具有分化成所有 3 种胚层的发育潜能。

对 ES 细胞的研究可以追溯到 50 年代对小鼠胚胎进行体外培养的实验。Whitten W.K.于 1956 年成功地应用补充了牛血清白蛋白的培养基培养受精卵,使单细胞的受精卵在体外发育到囊胚阶段。其后,Brinster R.L.于 1965 年建立了微滴培养技术,用于着床前小鼠胚胎的培养。Mclaren A.等对输卵管移植和子宫移植条件进行了优化,并最终使得通过胚胎移植从体外培养的胚胎产生活的小鼠成为一种常规的技术。10 年后,Gardner R.L.建立了将分离的细胞注射入宿主囊胚获得嵌和体小鼠的方法。这一方法的成功建立使得研究者有可能系统地检测经过体外操作的胚胎细胞的分化全能性。

1970 年,Stevens L.C.作为先驱者成功分离了小鼠畸胎瘤细胞(embryonal carcinoma, EC)并将其作为模式体系研究胚胎细胞的全能性。几年后,Brinster R.L.,Mintz B 和 Illmensee K.,以及 Papaioannou V.E.等证实胚胎瘤干细胞可以整合入宿主囊胚并分化成多种正常成年组织,但

没有胚胎瘤细胞整合入生殖系的报道。1981 年，Evans M.J.和 Marin M.H.分别独立地从正常的小鼠囊胚中分离了 ES 细胞。与大多数 EC 细胞不同，ES 细胞具有完全正常的核型。尤其重要的是，Bradley A.(1984)随后证实通过显微注射引入囊胚的 ES 细胞可以分化为成体的各种组织并整合入生殖系。小鼠 ES 细胞的分离和应用是 ES 细胞技术发展过程中的里程碑，它与同源重组技术的结合使得在生物整体水平上定向改变和修饰哺乳类动物的遗传物质成为可能。在此基础上发展起来的基因打靶技术革命性地改变了生命科学研究的面貌，成为研究基因功能的有力手段。

此后，陆续有从啮齿类动物之外物种的着床前胚胎中分离多潜能细胞的报道，但这些细胞的发育潜能没有得到进一步的研究和证实。1995 年，Thomson J.A.等报道从恒河猴中分离了猴的胚胎干细胞。这些细胞可以在体外培养 1 年以上并维持未分化的状态。这些细胞具有正常的核型，并保持分化成滋养层细胞、内胚层、中胚层和外胚层组织细胞的能力。猴 ES 细胞从细胞形态、细胞表面分子标记物、所需生长因子均不同于小鼠 ES 细胞，而与人类 EC 细胞类似。1996 年，Thomson J.A.等接着报道了绒猴 ES 细胞的分离。1998 年，Thomson J.A.等报道从人的囊胚中分离了 ES 细胞。这些人类 ES 细胞具有正常的核型，表达高水平的端粒酶活性。在体外不分化地增殖 4～5 个月后，这些细胞仍然保持分化成滋养层细胞和所有 3 种胚层的发育全能。Shamblott M.J.等在同年报道从流产胎儿的胚胎生殖细胞中成功分离了多能的胚胎干细胞。其实，早在 1994 年，Bongso A.等从人的囊胚中分离并培养了内细胞团来源的细胞，由于没有滋养层细胞的支持，这些细胞最终分化或者死亡。2000 年，他们报道了人 ES 细胞的建立，证实了 Thomson J.A.等人的研究并进一步对人 ES 细胞的表型及分化能力进行了鉴定。人源 ES 细胞的建立是一个激动人心的进展，对人 ES 细胞的研究将促进对人体早期发育、人类疾病的发生机制、新基因的发现、新药筛选和药效评价等基础和临床应用的研究。人源 ES 细胞的成功分离和培养以及体内和体外的定向分化研究，使得采用体外定向分化的组织细胞取代体内受损或衰竭组织细胞成为可能。随着这一领域的迅速发展，ES 细胞培养及定向分化技术正在成为新的细胞替代疗法的基础。

三、同源重组技术的发展

现代生命科学的研究由分子水平的研究向整体水平回归的趋势越来越明显。尤其是基因组计划完成之后，生命科学研究呈现出系统化、整体化和综合化的趋势。同源重组技术的发展也反应了这些特点。同源重组是指发生在减数分裂或者有丝分裂过程中相同或者相似 DNA 序列间的重组。它通常通过一对同源分子非姊妹染色体间的断裂而产生新的重组片段。同源重组技术的概念是 20 世纪 70 年代初在酵母研究中发展起来的。与酵母系统完全不同的是，哺乳动物细胞中同源重组的概率非常低。直到 1985 年，Smithies O.等才首次报道在肿瘤细胞中实现了人工打靶载体和内源 β－球蛋白基因间的同源重组。两年后，Smithies O.和 Capecchi M.R.的研究小组在小鼠胚胎干细胞中分别实现了次黄嘌呤磷酸转移酶基因(*Hprt*)的定位剔除。1988 年，Capecchi M.R.实验室的 Mansour S.L.等发展了一种称为“正负筛选”(positive－negative selection)的策略，将胸腺嘧啶激酶基因(*tk*)置于打靶载体的外侧，作为筛选的负性标记。与单用一种阳性筛选标记基因的策略相比，采用“正负筛选”策略时正确克隆富集的倍数要高 3～10 倍，真正使得同源重组技术可以作为修饰哺乳动物基因组遗传信息的常规手段并应用于生命科学研究的各个领域。

利用同源重组技术按照研究者的预先设计，在动物整体水平上灭活特定靶基因的研究开

始揭示出许多未知基因的功能。由于对基因组遗传信息进行精确修饰的需要，研究者随后发展了一些策略如"Hit and Run"以及"Tag and exchange"等，将基因的突变引入染色体组内。1994年，Gu H.等利用 Cre－LoxP 系统首次研制成功组织特异性基因剔除小鼠，使得对基因剔除进行时间和空间上的双向控制成为可能。1995年，Ramirez－Solis R.等应用位点特异的重组酶系统首次在小鼠 ES 细胞中实现了最长 3～4 cM 染色体组片段的缺失和倒位[31]。应用此策略可以构建染色体组任意片段的缺失突变体小鼠，用于遗传学分析，主要是用于鉴定缺失区域的隐性基因突变。近年来，世界上多个小鼠基因组中心纷纷采用大规模、系统的化学诱变进行突变小鼠的研究。2000年，de Angelis M. H.等报道了他们用乙烷亚硝基脲（N－ethyl－N－nitrosourea，ENU）大规模筛选基因显性突变的研究。他们对 14 000 只小鼠进行了临床相关指标的检测，获得了 182 只具有各种表型的突变小鼠。Nolan P. M.等在同时报道的研究中，从 26 000子代小鼠中分离到 500 种显性突变小鼠。将 ENU 诱变研究与染色体组缺失突变体研究结合起来，可以快速分析染色体组特定区域内的隐性突变，提高大规模基因功能研究的效率。

第二节　基因打靶技术在生物医学研究中的应用前景

一、基因打靶技术在基因功能研究中的应用

走进 21 世纪，生命科学已经进入了"后基因组"研究时代，大规模的基因功能研究正成为生命科学研究的热点。宏伟的人类基因组计划的顺利实施，完成了人类基因组及多种模式生物基因组的序列测定，产生了编码大约 30 000～40 000 个基因的核苷酸序列，其中只有一小部分基因的作用是已知的。面对这样一个打开的基因宝库，人类在新世纪面临的主要挑战是要解读这些序列所编码的基因的功能并研究基因调控的分子机制，这对于深入研究人类遗传疾病、肿瘤的发生机制以及理解许多生命现象的本质均有十分重要的意义。基因打靶技术的综合应用，为现代生物医学各个研究领域的高新技术及最新研究成果提供了一个研究系统，使得研究者可以在生物整体水平、组织水平、细胞水平、分子水平等各个层面上全面系统地研究基因功能。在基因功能研究方面，它具有其他方法所难以比拟的直接性和有效性，是"后基因组"时代进行功能基因组学研究的重要平台技术。从第一例成功的基因剔除报道至今，仅 13 年的时间，在包括美国、日本和欧洲各国在内的发达国家中，基因打靶技术已成为常规性的重要生物医学研究手段。在小鼠进行的基因打靶研究导致了生物医学研究领域中许多重要进展，小鼠已成为研究脊椎动物基因功能最重要的模式生物。利用基因打靶技术得到特定靶基因功能获得或者缺失的小鼠突变体并进行相关表型的研究将最终有助于绘制人类基因在各种生理和病理过程中功能的立体图卷。对基因功能的深入研究和了解将有利于发现新的具有应用价值的基因和蛋白，因此，通过基因打靶进行的基因功能研究将成为今后医药产业的基础。

基因打靶技术的发展使得它可以满足大规模基因功能研究的需要。目前，一种叫做"基因诱捕"的策略得到广泛的应用。利用基因诱捕可以建立一个携带随机插入突变的 ES 细胞库，节省大量筛选染色体组文库以及构建特异打靶载体的工作及费用，更有效和更迅速地进行小鼠染色体组的功能分析。此外，基因打靶研究结合随机化学物质（如 ENU）诱变的研究也正在世界上多个小鼠基因组中心进行。除了基因完全失活的小鼠模型，携带基因功能获得或者部分基因功能丧失突变的小鼠模型将会有助于对基因功能更全面和更深入的了解。

二、应用基因打靶技术研制人类疾病动物模型

随着生物医学研究的迅速发展，人们已经达成共识：所有的疾病，无论是器质性的还是功能性的，均可在基因水平上去探究其原因，并以此为基础研究早期诊断试剂，设计新的治疗方案，发展新的治疗药物。许多人类遗传病及肿瘤一旦确诊后已处于不可逆转的发生阶段，而且至今仍无理想的治疗药物。由于人类作为研究对象的局限性，发展相应的人类疾病动物模型，对于研究人类疾病的分子机制以及研制相应的诊断、治疗药物均具有非常重要的意义。

通过基因打靶研制的人类疾病小鼠模型已有数百种。这些模型的建立极大地丰富了人类对这些疾病相关基因功能的了解。以成纤维细胞生长因子受体 3(FGFR3)的功能研究为例，我们可以了解小鼠模型是如何帮助研究者发现该基因在长骨发育中的作用的。最早的临床研究发现 FGFR3 上不同功能域的突变导致不同程度的侏儒症。而在小鼠中进行的研究显示 Fgfr3 基因剔除导致与侏儒症病人完全相反的表型，*Fgfr3* 缺失突变导致小鼠长骨过度生长。这些数据提示 FGFR3 是长骨生长的负调控因子。此后，研究者利用基因敲入(gene knockin)和 Cre－LoxP 系统将 *Fgfr3* 突变引入小鼠基因组研制成功了侏儒症的模型小鼠，证实侏儒症确实是由 *Fgfr3* 突变导致基因功能激活所引起的。目前，这些模型小鼠正用于进一步的功能研究和 *Fgfr3* 抑制物的筛选，研究者希望通过药物筛选可以获得 *Fgfr3* 活性的抑制物以用于这种遗传疾病的治疗。

许多肿瘤病人都有家族史，家族中的人对某种特定的肿瘤表现出一种易患性。现在知道这种家族易感性与肿瘤抑制基因的突变经生殖系遗传有关。用基因打靶技术已经研制出了数十种小鼠肿瘤模型，其中最著名的当属肿瘤抑制基因 *P53* 基因剔除小鼠。*P53* 基因突变导致人类 Li－Fraumeni 综合症，患者并发各种类型的肿瘤，主要是软组织肉瘤、乳腺癌、脑瘤、骨瘤、白血病等。*P53* 基因剔除小鼠精确地模仿 Li－Fraumeni 综合症患者的表型，其肿瘤表型受小鼠种系遗传背景的影响。因此，通过对这些小鼠的研究，可能发现 *P53* 基因相关的影响肿瘤发生率的修饰基因。由于肿瘤抑制基因往往在生命的基本活动中担当重要的功能，许多肿瘤抑制基因如 *Brca*1、*Dpc*4/*Smad*4 定位缺失导致小鼠在胚胎期死亡，使得研究者无法研究这些基因在成体器官发育和肿瘤发生中的功能。现代基因打靶技术的发展已使得对基因打靶进行时间和空间上的控制成为可能，研究者利用条件基因剔除技术建立了多种组织特异的肿瘤模型。如乳腺癌肿瘤抑制基因 *Brca*1 基因剔除导致小鼠在中胚层发生前死亡。研究者利用 Cre－LoxP 系统研制了乳腺组织特异性 *Brca*1 基因剔除小鼠。经过一个很长的潜伏期之后，突变小鼠长了乳腺癌，在组织病理学上与人类的乳腺癌惊人的相似。对该小鼠进行表型分析的结果表明，*Brca*1 基因缺失导致基因组不稳定性增加，引起了其他基因包括肿瘤抑制基因和癌基因的突变，最终导致乳腺癌发生。这些小鼠除了可被继续用于研究相关疾病的分子机制，也可用于治疗药物的筛选，并可用作传统治疗药物和基因治疗的评价系统。

三、应用基因打靶技术改良动物品系和研制动物反应器

自 1980 年第一例转基因小鼠诞生以来，通过显微注射法将外源基因片段引入受精卵的雄性原核一直是研制转基因动物最实用的方法。采用这种传统方法获得的转基因动物，其外源基因整合的位点和拷贝数是高度不确定的，对外源基因的表达以及由此而引起的表型有很大的影响。传统转基因载体上要设法带上所有的调控元件以期能使外源基因达到理想的表达。许多时候，某一基因的调控元件在染色体组上相距甚远，或存在于内含子中，使得外源基因要

得到理想的表达十分困难。应用以 ES 细胞培养技术和同源重组为基础的基因敲入技术，可将外源基因整合到特定的靶位点，克服外源基因整合的位点效应，并可利用靶位点全套的表达调控元件以实现组织与时空特异性的表达。由于经济大动物的胚胎干细胞系还鲜有报道，克隆技术却有突飞猛进的发展，在体细胞中进行基因打靶并通过核移植获得转基因动物是一条可行的途径。2000 年，McCreath K.J. 等成功地利用基因打靶技术在绵羊胎儿成纤维细胞中将 α 抗胰蛋白酶(AAT)基因定位整合入绵羊(I 原胶原，COL1A1)基因位点，并通过核移植技术获得了一只活的绵羊，外源基因在乳汁中的含量达 0.65 mg/mL。显而易见，通过基因打靶技术研制动物生物反应器正在改变传统研制转基因动物的方法。此外，由于高级哺乳动物恒河猴和人的 ES 细胞相继成功地建立，有理由期待其他经济大动物的 ES 细胞在近期成功建立。随着对人类基因组功能了解的逐渐加深，人类对高等动物遗传物质的改造能力也将逐渐增强，利用基因打靶技术定向改变这些经济大动物的经济性状是完全有可能的。

第三节　研制基因打靶小鼠的实例——*Smad5* 基因剔除小鼠的研制和鉴定

一、靶基因 cDNA 探针的克隆

进行基因打靶研究的第一步是要获得靶基因的 cDNA 片段。多聚酶链式反应(PCR)因其简便、快速、敏感、特异等特点成为首选方法。选择靶基因相对富集的 cDNA 文库作为模板，或者从特定组织提取 mRNA，通过逆转录反应合成 cDNA 的第一条链作为 PCR 反应的模板。若是已知基因，可根据基因序列合成特异性的引物；对未知基因，可根据蛋白质序列或同源基因序列合成简并引物。对表达水平较低的基因，针对 PCR 产物合成另一对内源的引物进行巢式或半巢式 PCR，将提高成功的概率。图 10－2 显示采用不同 PCR 引物从小鼠胰腺 cDNA 文库和 RI－6 细胞 cDNA 文库中扩增到 *Smad5* 的 cDNA 片段。利用 TA 克隆法将 PCR 产物克隆，克隆的 cDNA 片段可用作进一步克隆靶基因基因组 DNA 片段的探针。

二、靶基因基因组 DNA 片断的筛选和克隆

可以通过筛选 129 小鼠基因组噬菌体文库或者细菌人工染色体文库来获得靶基因基因组 DNA 片段。129 小鼠基因组细菌人工染色体(BAC)文库(genome system, cat. No FBAC 4432)由 5 张尼龙膜组成，每张膜上点满了高密度的 BAC 克隆，每个 BAC 克隆包含有平均长度为 100 kb 的基因组片段。用同位素标记靶基因探针，并参照公司提供的实验手册进行文库筛选，最后从该公司购买含有靶基因的阳性克隆。每个克隆含有包括靶基因在内的大约 100 kb 基因组片段。

将含有 *Smad5* 基因组片段 BAC 质粒的细菌在含有氯霉素的 LB 培养基中生长，并用 SDS 裂解法提取质粒。将提取的 DNA 用多种限制性内切酶消化。图 10－3A 所示为含有 *Smad5* 基因组片段的 BAC 质粒用不同限制性内切酶消化的结果。将 DNA 转移到硝酸纤维素膜上并用 *Smad5* 基因探针进行杂交反应，结果如图 3B 所示。由于 BAC 质粒是氯霉素抗性的，所以可以采用鸟枪法将靶基因片段克隆到氨苄青霉素抗性的克隆载体 Bluescript SK 上。通过杂交实验可以帮助选择合适的限制性内切酶进行亚克隆。从图 10－3B 结果可以看出 *Xba* I、*Hind* III、*Sac* I 在靶基因上均有多处识别位点，不宜用于制备亚克隆。*Sal* I 片段太大，超过 23 kb，克隆到普通的质粒上也有困难。*EcoR* I 片段约为 9 kb，用于载体构建并筛选合适的 Southern 探针

略显不足。*BamH* I 片段大小介于 9～23 kb 之间，这样大小的片段既便于克隆，又使得选择同源臂和探针有较大的余地。因此，最终将 *Smad5* 基因组克隆用 *BamH* I 消化后经鸟枪法克隆到克隆载体 Bluescript SK 上，获得了 19 kb 的 *Smad5* 基因组片段。

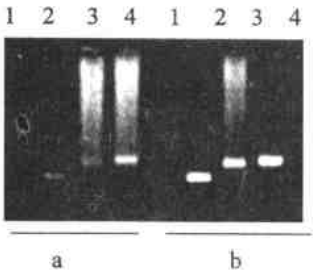


图 10－2　PCR 法获得小鼠 *Smad5* 基因 cDNA 片段

1～4 为 4 对不同的引物。A，模板为小鼠胰腺 cDNA 文库；B，模板为小鼠 R1－6 细胞 cDNA 文库。

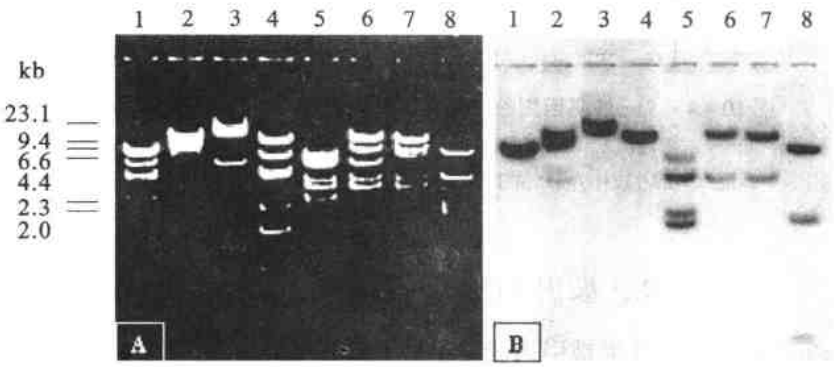


图 10－3　含有 *Smad5* 基因的 BAC 质粒的杂交分析。

A. 含有 *Smad5* 基因的 BAC 质粒的限制性内切酶分析结果。B. 用 *Smad5* cDNA 探针进行 Southern 杂交的结果。1，*EcoR* Ⅰ；2，*Xba* Ⅰ；3，*Sal* Ⅰ；4，*Bam H* Ⅰ；5，*Hind* Ⅲ；6，*EcoR* VICla Ⅰ；7，*Sac* Ⅱ/*EcoR* V；8，*Sac* Ⅰ

三、基因剔除打靶载体的构建以及中靶 ES 细胞的筛选

经酶切，将 12 kb 的 *Smad5* 基因组片段用于构建打靶载体，分成长臂和短臂插入打靶载体 pLoxpneo[54] 中，如图 10－4A 所示。阳性筛选标记基因被插入到 *Smad5* 基因的第 6 外显子中，阴性筛选标记基因位于同源臂长臂的外侧。通常，打靶载体同源臂的总长度应大于 6 kb，短臂长度应大于 2 kb。打靶载体通过电穿孔转染小鼠胚胎干细胞（ES 细胞），在含有 G418/FIAU 的培养基中筛选双抗性的 ES 克隆。从阳性克隆细胞中制备基因组 DNA 进行 Southern 杂交以筛选发生了正确同源重组事件的中靶 ES 细胞，部分结果如图 10－4B 所示。基因组 DNA 用 *Asp*718 和 *BamH*I 进行消化，利用打靶载体同源臂外侧的 DNA 作为探针进行 Southern 杂交，野生型等位基因将产生一条 10 kb 的阳性条带，中靶等位基因由于 *neo* 基因的插入将显示 12 kb 的阳性条带。

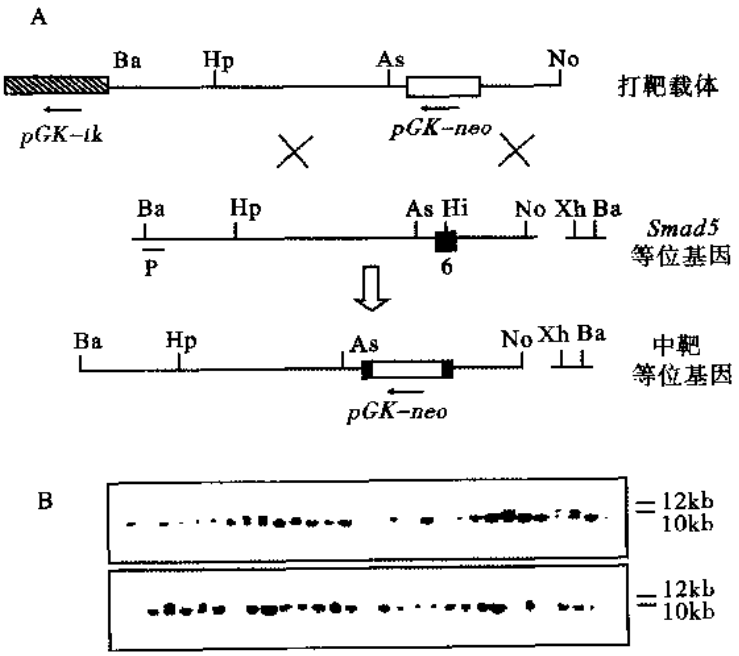


图 10-4　*Samd5* 基因剔除打靶载体的构建以及中靶 ES 细胞的筛选。

A.构建打靶载体的策略。黑色正方形表示第 6 外显子。B.Southern 杂交筛选中靶 ES 细胞。基因组 DNA 用 *Bam HI* 和 *Xho* I 消化，用 A 中标记为 P 的片段作为杂交的探针。

四、基因剔除小鼠的获得及鉴定

中靶 ES 细胞通过显微注射被引入受体小鼠 C57BL/6J 的囊胚。注射后的囊胚通过胚胎移植引入假孕母鼠的子宫中并使其发育至出生。获得的高嵌合度子代小鼠与野生型小鼠杂交以获得经种系遗传的 *Smad5* 基因剔除杂合子小鼠。将 *Smad5* 基因剔除杂合子小鼠进行杂交可获得 *Smad5* 基因剔除纯合子小鼠。对 176 只 *Smad5* 基因剔除杂合子小鼠交配产生的后代进行基因型鉴定的结果(表 10-1)显示没有纯合子小鼠，表明 *Smad5* 基因剔除导致小鼠胚胎期死亡。为了了解 *Smad5* 纯合子胚胎死亡的时间，对不同时期的小鼠胚胎进行了表型观察和基因型鉴定(表 10-1)，结果表明，异常发育的胚胎在胚胎期 E8.5-E11.5 出现，*Smad5* 纯合子胚胎在 E10.5-11.5 死亡。所有这些实验证据表明，*Smad5* 基因缺失导致小鼠在胚胎期 E11.5 前死亡。*Smad5* 基因剔除突变基因可以通过杂合子小鼠稳定遗传。

对于任何基因打靶突变型小鼠，一般可用 3 条寡核苷酸配成 2 对 PCR 反应引物，通过 2 个 PCR 反应区分野生型、杂合型和纯合型等 3 种不同的基因型。在打靶载体 *neo* 基因插入位点两侧设计引物 1(5′-TAAACAATCGTGTTGGGGAAGC-3′)和引物 2(5′-AATGAATTGTATCCTGCCGGTA-3′)，两者能从野生型 *Smad5* 等位基因上扩增出 231 bp 的阳性条带。如果用引物 2 和 *neo* 基因上的引物 3 配对进行 PCR 反应，则可从突变等位基因上扩增出 290 bp 的阳性条带。为了更有效地区分野生型等位基因和突变型等位基因，本实验采用引物 3 和引物 4 配对进行 PCR 反应(图 10-5A)，从突变等位基因上扩增出 510 bp 的条带。如图 10-5B 所示，来自每个胚胎的基因组 DNA 分别用 P1/P2 以及 P3/P4 进行 PCR 反应，只能扩增出 231 bp 条带的

胚胎为野生型；同时扩增出 231 bp 和 510 bp 条带的胚胎为杂合型；只能扩增出 510 bp 条带的胚胎为突变纯合型。

利用来自野生型和纯合子胚胎的 RNA 进行 Northern 杂交，结果如图 10－5C 所示，*Smad5* 突变纯合子胚胎中检测不到任何 *Smad5* 基因的转录本，而与 *Smad5* 高度同源的 *Smad1* 基因在野生型胚胎和突变型胚胎中均正常表达。这些结果表明，采用以上策略进行 *Smad5* 基因剔除导致无义突变(null mutation)。

五、基因剔除小鼠的表型分析

利用组织和病理学分析以及分子生物学技术对基因剔除小鼠进行表型分析，可以了解靶基因在哺乳动物发育过程中的功能。*Smad5* 纯合子胚胎在 E11.5 前死于多种胚胎及胚外组织缺陷(图 10－5D)，表明 *Smad5* 基因是小鼠胚胎正常发育所必需的。*Smad5* 突变胚胎在 E9.5 后表现出与 BMP－4、TGF－I 及 TGF－II 型受体基因剔除小鼠类似的血管形成缺陷的表型。*Smad5* 缺陷型胚胎的血管形成可以发生，因为其卵黄膜及胚胎均可检测到内皮细胞粘附分子(pECAM)阳性的血管内皮细胞。进一步的研究结果表明 *Smad5* 突变胚胎血管直径增加，不能形成正常的血管网络，可能与胚胎内增加的间充质细胞凋亡以及血管外周减少的平滑肌细胞有关。这些结果显示，*Smad5* 可能介导 TGF－I 信号在毛细血管形成过程中调节内皮细胞－间充质细胞间的相互关系。*Smad5* 在原肠期发生的 E7.0 天便在小鼠胚胎表达。它在胚胎期 E9.5 及 E10.5 天的表达与 TGF－I 及 TGF－II 型受体的表达高度一致。值得注意的是，以前的

表 10－1　*Smad5* 杂合子小鼠交配产生子代小鼠的表型和基因型分析

年龄	正常	表型不正常	吸收	基因型			总计
				+/+	+/−	−/−	
E8.5	35	2	1	9	18	20	38
E9.5	79	22	2	27	52	22	103
E10.5	48	13	2	18	30	15	63
E11.5	40	2	5	13	27	17	57
出生后	176	0	0	71	105	0	176

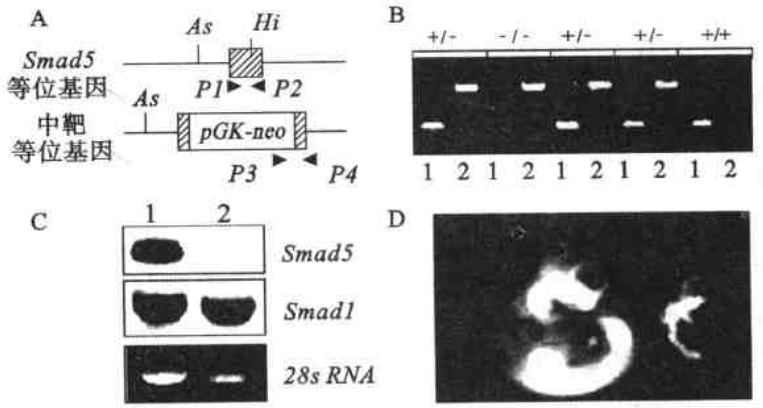


图 10－5　*Smad5* 基因剔除小鼠的基因鉴定及其表型

A.基因型鉴定引物的设计。P1、P2 和 P4 位于 *Smad5* 基因的第 6 外显子中，P3 位于阳性筛选标记基 pGK－*neo* 中。B.PCR 鉴定基因型结果。1. P1/P2 扩增 *Smad5* 野生型等位基因；2. P3/P4 扩增 *Smad5* 基因剔除等位基因。C.Northern 杂交结果。1. 野生型胚胎；2. *Smad5* 突变纯合子胚胎。D.E10.5 野生型胚胎(左)和 *Smad5* 突变纯合子胚胎。

许多体外实验结果将 *Smad*5 置于 BMP 信号转导通路而非 TGF－I 通路中，*Smad*5 基因剔除小鼠与 TGF－I 及 TGF－II 型受体基因剔除小鼠相似的表型，提示体内各 *Smads* 信号通路可能存在复杂的相互作用。

第四节　常用实验操作方法

一、小鼠基因组 DNA 的制备

剪下出生后 10～14 d 小鼠的尾尖约 0.5 cm，置于 1.5 mL 离心管中，加入 400 μL 组织裂解缓冲液(0.5% SDS，0.1mol/L NaCl，0.05 mmol/L EDTA，0.01 mmol/L Tris.Cl，pH8.0，蛋白酶 K 100 μg/mL)，50 ℃保温过夜，再按以下方法制备基因组 DNA。

1. 每离心管加入 200 μL 饱和的 NaCl(6 mol/L)。
2. 将离心管置于纸盒中剧烈晃动 200 次。
3. 置于冰上，10 min。
4. 室温 14 000 r/min 离心，10 min。
5. 将上清 500 μL 转移至干净的 1.5 mL 离心管中，每管加入 0.8 mL 乙醇，混匀。
6. 14 000 r/min 离心，5 min。
7. 室温干燥，将 DNA 重悬于 50～100 μL TE 中，37 ℃保温至 DNA 完全溶解后可用于 PCR 或 Southern 杂交分析。

二、基因组 DNA 的 PCR 分析

1. 用于 PCR 反应的基因组 DNA 沉淀可于 50 ℃溶解 2 h。设置 PCR 反应前将样品振荡混匀。若样品将用于 Southern 杂交分析，则最好将样品置于 37 ℃溶解过夜，并避免振荡。

2. 设置 PCR 反应：

待测 DNA 样品	1.0 μL
5′引物 (20 pmol/ μL)	1.0 μL
3′引物 (20 pmol/ μL)	1.0 μL
20 mM dNTP	0.125 μL
Taq 聚合酶	0.125 μL
10×PCR 反应缓冲液	2.5 μL
H_2O	19.25 μL

3. PCR 反应条件：

94 ℃，10 min
94 ℃，2 min
94 ℃，30 s；60 ℃，30 s；72 ℃，2 min，30 个循环。
72 ℃，10 min

4. 在 2% 琼脂糖凝胶上电泳并用溴化乙锭染色。

三、基因组 DNA 的 Southern 杂交分析

1. 用一种或多种限制性内切酶消化适量的基因组 DNA(约 10 μg DNA 于 30 μL 反应体

积)。

2. 在琼脂糖凝胶(0.65% ~0.8%)上缓慢电泳(22 V), 24~48 h。

3. 电泳完毕后在紫外光下照相,沿凝胶边缘放置一透明荧光直尺,以便能从照片中读出DNA 标准参照物的迁移距离。

4. 将凝胶置于数倍体积的 0.25 N HCl 中浸泡 20 min,并且温和地不断振摇。

5. 将凝胶浸泡于数倍体积的变性缓冲液中(1.5 M NaCl,0.5 N NaOH),浸泡 30 min,温和振摇。

6. 弃去变性缓冲液,加入数倍体积的中和缓冲液(1 M Tris·HCl pH7.4, 1.5 M NaCl),于室温不断振摇 30 min。

7. 将凝胶置于 20 × SSPE 中,室温 30 min。

8. 用毛细管转移法或电转移法将 DNA 中琼脂糖凝胶电脉转移到硝酸纤维素滤膜上。

9. 转移结束后,用铅笔标记凝胶加样孔的位置。

10. UV 照射交联(120 mg/cm^2)。

11. 2 × SSPE,室温漂洗。

12. 将滤膜置于两组 3 MM 滤纸中间,用真空炉于 80 ℃干烤 1 h。

13. 将滤膜置于预杂交液中(10 × Denhardt's,4 × SET,0.1% SDS,0.1% $Na_2H_2P_2O_7$,100 μg/mL 变性的鲑精 DNA), 杂交瓶于 65 ℃温育 1 h。

14. 将 DNA 探针和鲑精 DNA,于 100 ℃加热 5 min 使其变性,迅速置于冰上 5 min。

15. 取 200 μL DNA 探针和 100 μL(10 mg/mL)鲑精 DNA 加入 10 mL 新的预杂交液中。

16. 将杂交瓶中的预杂交液弃去,加入含有变性探针的杂交液。65 ℃杂交过夜。

17. 将滤膜转移至盛有数百毫升的漂洗缓冲液(0.4 × SET,0.1% SDS,0.1% $Na_2H_2P_2O_7$)中,于 65 ℃水浴摇床中温和摇动漂洗 2 次,每次 10min。

18. 将滤膜置于两张 3 MM 纸中稍事干燥,用 Saran 保鲜膜包好滤膜,贴上数张荧光标签,以便以后校准放射自显影与滤膜的位置。

19. 将滤膜置于 X 光片夹中于 -70 ℃加增感屏曝光 12~48 h。

20 × SET 缓冲液:

3 M	NaCl
0.4 M	Tris·HCl, pH7.5
20 mM	EDTA

四、用原代小鼠胚胎成纤维细胞制备饲养层

具有分化全能性的小鼠胚胎干细胞(ES 细胞)长时间的培养和维持必须利用成纤维细胞作为饲养层,或是在添加了 LIF 的培养基或条件培养基中生长。从各种小鼠品系中分离的胚胎成纤维细胞在支持 ES 细胞生长的功能方面没有显著性差异。如果胚胎成纤维细胞要用于打靶实验,还必须具有 G418 的抗性,通常可从携带 *neo* 基因的转基因小鼠或者基因剔除小鼠中分离。以下是分离胚胎成纤维细胞的具体方法:

1. 将怀孕 14 d 的母鼠处死,打开腹腔,取出带有胚胎的子宫。在 PBS 中将胚胎从子宫中解剖出来,放入无菌的 100 mm 培养皿中。

2. 去除卵黄囊、羊膜和胎盘,将胚胎在新鲜的 PBS 中洗 2 次。

3. 用一对镊子将胎盘的头及内脏去除,用消毒剪刀将胎鼠剪成小的碎片(1~3 mm^3),加

入 PBS 洗至液体基本无色。

4. 将胎鼠碎片转移至含 PBS 的 15 mL 离心管,1 000 r/min,离心 2 min。

5. 弃上清,加入 5 mL 0.25% 胰蛋白酶,37 ℃,处理 30 min,加入含 10% 胎牛血清的 DMEM 培养基,用吸管反复吹打。

6. 1 000 r/min, 离心 2 min。

7. 弃上清,加入饲养层细胞培养基并将细胞转移到 150 mm 培养皿中置于含 5% CO_2的细胞培养箱中 37 ℃培养 2~3 d,通常第 3 d 胚胎成纤维细胞已长满培养皿。

8. 将胚胎成纤维细胞从 1 个培养皿转移至 3 个培养皿,3 d 后可将细胞用含 10% DMSO 的培养层细胞培养基冻存。这种细胞即是原始小鼠胚胎成纤维细胞。

原始小鼠胚胎成纤维细胞用作饲养层细胞之前,必须用丝裂霉素 C(mitomycin C)按以下方法处理,使它们停止生长,这样它们既能支持 ES 细胞的生长又不会比培养的 ES 细胞生长的快。

1. 取 1 管原始小鼠胚胎成纤维细胞在 40 ℃水浴快速解冻,将细胞转移至有 2 mL 培养基的离心管中,1 000 r/min,离心 2 min。

2. 弃上清,加入新鲜的饲养层细胞培养基,并转移至 2 个 100 mm 的培养皿中,置于 37 ℃的 CO_2培养箱中培养。

3. 3 天后,将细胞用胰蛋白酶消化并转移至 6 个 150 mm 培养皿中,37 ℃培养 3 d。

4. 将细胞从 6 个 150 mm 培养皿转移至 12 个 150 mm 培养皿中,37 ℃培养 3~4 d。

5. 将细胞从 12 个 150 mm 培养皿中转移至 40 个 150 mm 培养皿中,37 ℃培养 3~4 d。

6. 弃去培养基,用 PBS 洗一次。向每个培养皿中加入 3 mL 0.25% 胰蛋白酶。处理 10 min 后,加入 5 mL 饲养层细胞培养基终止反应。

7. 将细胞收集入 50 mL 离心管,1 000 r/min,离心 2 min。

8. 加入冻存培养基(15% FCS, 10% DMSO,75% DMEM), 总共 40 mL,用吸管反复吹打混匀细胞,迅速分成 40 小管冻存于 -80 ℃备用。这样制备的饲养层成纤维细胞每管解冻后通常可以铺满 3~4 个 90 mm 培养皿。

准备 600 mL 饲养层细胞培养基可参考以下配方:

500 mL	DMEM
6 mL	L-谷氨酰胺
6 mL	非必需氨基酸溶液(10 mM, 100×)
6 mL	青链霉素溶液(5 000 单位青霉素, 5 mg 链霉素/100 mL)
4 ml	巯基乙醇
90 mL	胎牛血清

用一次性过滤瓶过滤,置于 4 ℃冰箱中备用。在此培养基中添加 1 000 U/mL LIF 即可用于 ES 细胞的培养。

五、小鼠胚胎干细胞的常规培养

小鼠胚胎干细胞的培养和遗传操作是基因打靶过程中最关键的环节。这一环节之所以重要,是因为必须在 ES 细胞培养过程中维持其正常核型,并保证在被引入受体囊胚时保持其高效形成功能性生殖细胞的能力。另外这一阶段的实验费用昂贵、工作量大、实验周期长。妥善维持转染前的 ES 细胞和中靶后的 ES 细胞均同样重要。由于 ES 细胞对去垢剂高度敏感,故建

议在 ES 细胞培养中尽量使用一次性培养器皿。正常的 ES 细胞在饲养层细胞上应形成隆起的、致密的、边缘光滑的集落。如果 ES 细胞克隆变得扁平或者表面变为粗糙的内胚层细胞，在使用这些 ES 细胞进行打靶实验前，应重新检测这种 ES 细胞整合到生殖系的能力。

不同的 ES 细胞系对在体外培养并保持未分化状态的生长条件要求不尽相同。由于建系之初所用的培养条件不同，因此大多数 ES 细胞在生长培养基、传代之间的生长时间、每一代所接种的细胞数、胰蛋白酶消化的程度、培养箱中的 CO_2 浓度、饲养层细胞的类型都不尽相同。另有几个在培养过程中应注意的细节是：ES 细胞应以较高的密度培养，每 2～3 d 1∶3 传代一次，每天更换培养基；成团的 ES 细胞容易分化，因此每次用胰蛋白酶消化 ES 克隆时，应力争使 ES 细胞全部变成单细胞；应尽量减少 ES 细胞在体外的操作时间，同时尽可能多地冻存早期的 ES 细胞或者中靶 ES 细胞；用于打靶实验或者囊胚显微注射的 ES 细胞在实验前不能在体外培养时间太长。

一般 ES 细胞每 2～3 d 以 1∶3 比例传代一次。培养 ES 细胞的培养皿在接种饲养层细胞前最好用 0.1% 明胶处理。

1. 弃去原培养基，加入 1.5 mL PBS 洗 1 次。

2. 加入 1.5 mL 0.25% 胰蛋白酶－EDTA 溶液，消化 5 min 后，加入含 15% 胎牛血清的 DMEM 高糖培养基中和胰蛋白酶。

3. 用吸管轻轻吹打，直至成为单细胞悬液。

4. 将细胞悬液按 1∶3 的比例转移到新的含有饲养层的 35 mm 细胞培养皿中，补加一些相同的培养基。

5. 将细胞悬液摇匀后，置入 5% CO_2、37 ℃培养箱中培养。

6. 每日换液一次，若细胞长满可传代。

六、ES 细胞的冻存和复苏

ES 细胞可以象其他组织培养细胞一样，用 10% DMSO 作为冻存保护剂进行冻存。与常规方法一样，细胞应慢冻速融。对于长期保存，细胞应冻存于液氮；对于短期保存(6 个月内)，细胞可以保存于 －70 ℃冰箱。ES 细胞的最佳冻存密度为 $5\times10^6\sim10\times10^6$ cells/mL。每个 100 mm 培养皿培养的细胞约可冻存于 3 个 1 mL 的冻存管。复苏细胞时要快，尽可能去除培养基中的 DMSO。

参照以下方法冻存 ES 细胞：

1. 以一个 100 mm 组织培养皿中的 ES 细胞为例。首先吸去培养基，加入 10 mL PBS。

2. 吸去 PBS，加入 1 mL 0.25% 胰蛋白酶－EDTA，37 ℃消化 5 min。

3. 加入 5 mL 含 15% 胎牛血清的饲养层细胞培养基，用吸管反复吹打(至少上下 20 次)，使 ES 细胞尽可能地分散成单细胞。

4. 收集细胞悬液于 15 mL 离心管中，1 000 r/min，离心 5 min。去除上清，重新悬浮细胞于 1.5 mL饲养层细胞培养基中。

5. 分装 ES 细胞悬液于 1 mL 冻存管中，每管 0.5 mL。冻存管中预装 0.5 mL 细胞冻存液(20% DMSO，25% FCS，55% DMEM，每次新鲜配制)。

6. 混匀每管细胞，快速转移到预冷的泡沫塑料盒中，保存于 －70 ℃冰箱。

7. 24 h 后转移冻存细胞至液氮中。

参照以下方法复苏 ES 细胞：

1. 取出冻存管，在预热至 37～39 ℃的水中迅速摇动，使细胞悬液尽快融化。

2. 将细胞悬液由冻存管转移至一无菌离心管中，离心管中预先加入了 3 倍体积的含 15% 胎牛血清的 DMEM 高糖培养基，混匀。

3. 离心管加盖离心，1 000 r/min，10 min。

4. 弃上清，加入常规培养 ES 细胞的培养基，悬浮成单细胞悬液。

5. 将细胞悬液移入接种有饲养层的 100 mm 培养皿中，补加适量的培养基至总体积约为 12 mL。

6. 培养皿放入 5% CO_2，37 ℃培养箱中培养，次日换液。长满后，可按常规方法传代。

（杨　晓）

第十一章

转基因动物技术

内容提要

转基因动物的原理和方法中主要介绍了有关转基因概念、目的基因的设计和两种转基因方法的原理。以大鼠为例，介绍显微注射法进行转基因的操作步骤，包括仪器试剂、外源基因和实验动物准备和显微注射技术。简要介绍转基因整合与表达的检测方法。

转基因动物自 1980～1981 年产生后，迅速发展。它克服了物种之间的生殖隔离，实现了动物物种之间遗传物质的交换和重组，这不仅为遗传物质的研究提供了新的手段，丰富了物种的基因库，也扩大了生命科学的视野。由于转基因动物技术具有分子及细胞水平操作、整体表达的特点，使得人们能从更完整的系统中去认识生物分子的作用，现已广泛应用于医药、农业及生物学领域。在医学领域已用于遗传病、肿瘤、心血管疾病、免疫性疾病的研究，包括建立疾病的病理模型，探讨其病因、发病机理和基因治疗方法，研究基因表达调控以及进行药物干预和新药的开发实验等。建立疾病的转基因动物模型是转基因动物技术应用于医学方面的一个重要研究内容。转基因技术可以在动物原来遗传背景的基础上，通过改变某种基因的表达水平以建立人类疾病的动物模型。这种模型产生疾病的原因清楚(由转入的外源基因引起)，模型动物的症状单一，接近于病人的症状。在人类遗传病研究方面对单基因遗传病，已成功地建立了镰刀状细胞贫血、囊性纤维化、家族性高胆固醇血症和腺苷脱氨酶缺乏症等的转基因动物模型，以研究单个基因在遗传病发病中的作用。在农业方面，转基因农业动物研究有了很大进展，尤其在通过胚胎细胞或体细胞将基因导入牛、猪、羊和山羊，产生可遗传的转基因牲畜方面发展很快。说明转基因牲畜不仅可作为生物反应器，为人类生产基因工程产品和用于牲畜品种改良，提供更好的服务，同时也促进了胚胎学和发育遗传学的飞速发展，并开拓了医学研究的新领域。

第一节　原理和方法

一、转基因、转基因动物与转基因动物模型

转基因动物(transgenic animals)是指基因组中整合有外源基因的一类动物；整入动物基因组的外源基因被称为转基因(transgene)。由于建立转基因动物时，外源基因可能只整合入动物

的部分组织细胞的基因组，也可能整合进动物所有组织细胞的基因组中。我们把只有部分组织的基因组中整合有外源基因的动物，称为嵌合体动物（chimera，mosaic animals），这类动物只有当外源基因整合进去的"部分组织细胞"恰为生殖细胞时，才能将其携带的外源基因遗传给子代；否则，外源基因将不能传给子代。一般用胚胎干细胞法，逆转录病毒载体法产生的第一代转基因动物均为嵌合体动物，而显微注射法得到的第一代转基因动物，也约有 20% 为此类动物。如果动物所有的细胞都整合有外源基因，则具有将外源基因遗传给子代的能力，通常把这类动物称为转基因动物。当外源基因在转基因动物体内表达，并培育出其表型与人类疾病症状相似的动物模型，则称其为转基因动物模型。

二、目的基因的设计

目前建立转基因动物的主要策略有两种。一种是让转基因在动物体内过度表达的方法。最常用的是显微注射方法。转基因可用基因本身的启动子，也可拼接组织特异性表达的外源启动子；可转入单基因，也可转入双基因。另一种方法是让基因在体内灭活，丧失其功能。这就是近年来发展的用胚胎干细胞进行基因打靶技术，以产生基因缺陷的转基因动物。目前这种方法获得成功的多半是在小鼠上进行的。

这里所指的目的基因即所说的外源基因。首先要针对实验目的设计外源基因，要么设计随机整合型基因，要么设计基因敲除（knockout）型载体基因。用显微注射方法进行转基因动物时，随机整合型基因从结构上看可分为两部分，即包括含调控元件（regulatory element）的旁侧序列（flanking sequence）和可表达的结构基因（structural gene）序列。调控元件部分：①选用结构基因的天然启动子，其基因必须完整，应包括 5′端上游的启动子（promoter）区和 3′端下游的加尾信号（poly（A）signal）区等。一般所含周围序列长度在 500 bp 范围内，即含有决定基因组织特异性的调控元件。以转基因自己的调控元件进行的整合可增加转基因动物基因的拷贝数，以产生组织特异性高表达。②选用具有高表达活性的强启动子，使基因在特定的组织和器官过度表达。一般情况下，需先删除结构基因的天然启动子，将来自于不同基因的强启动子序列（甚至增强子序列）和结构基因拼接成融合基因使其高表达，或同时具有组织特异性高表达特性。结构基因部分可选用基因组 DNA 片段或基因的 cDNA，后者因不含有内含子可使外源基因有可能不表达或弱表达。为研究基因的调控元件，有时结构基因部分可选用易于检测的报告基因。通常使用报告基因是由于转基因与内源基因的同源性较高，使转基因不易检测，为此常常在结构基因的内含子中插入或在其后随意拼接一个长度为几十至几百碱基的报告基因（又称报告序列）。用基因敲除进行转基因动物时，首先要设计和待敲除基因有同源性的靶基因载体即打靶载体，包括用于正负筛选的标志基因，利用同源重组原理进行同源代替或同源插入，从而使目的基因失活，故可用于建立基因缺陷动物。

三、转基因的方法

转基因动物的基本原理是将改建后的目的基因用显微注射等方法注入实验动物的受精卵（或着床前的胚胎细胞），然后再将其植入受体动物的输卵管（或子宫）中，使其发育成携带有外源基因的转基因动物，（或先为嵌合体动物，再培育为转基因动物）。人们可以通过分析转基因和动物表型的关系，从而揭示外源基因的功能，培育品种优良的工程动物或具有与人类疾病症状相似表型的动物模型。以下介绍最常用的也是比较成功的两种方法。

(一)显微注射法

自从 1980 年 Gorden 等首次将显微注射技术(microinjection technique)用于小鼠受精卵的基因导入以来,该技术已得到广泛应用。这种技术多数是用小鼠,现已用于大鼠、兔、山羊、绵羊、牛和猪。在心血管研究中发展的转基因大鼠技术具有特殊的重要意义,这是因小鼠体积小,取材受限。显微注射技术是最常用的转基因方法之一。这种技术采用显微注射仪对实验动物的受精卵原核进行注射外源基因并使其整合到动物的基因组,从而得到转基因动物。用显微注射产生转基因大、小鼠的实验步骤如图 11－1 所示。约有 20%的大、小鼠的子代可整合有外源基因。外源基因整合的拷贝数从一至数百个不等,一般情况下在染色体的单个位点整合,有时可能在两个不同染色体的不同位点整合。

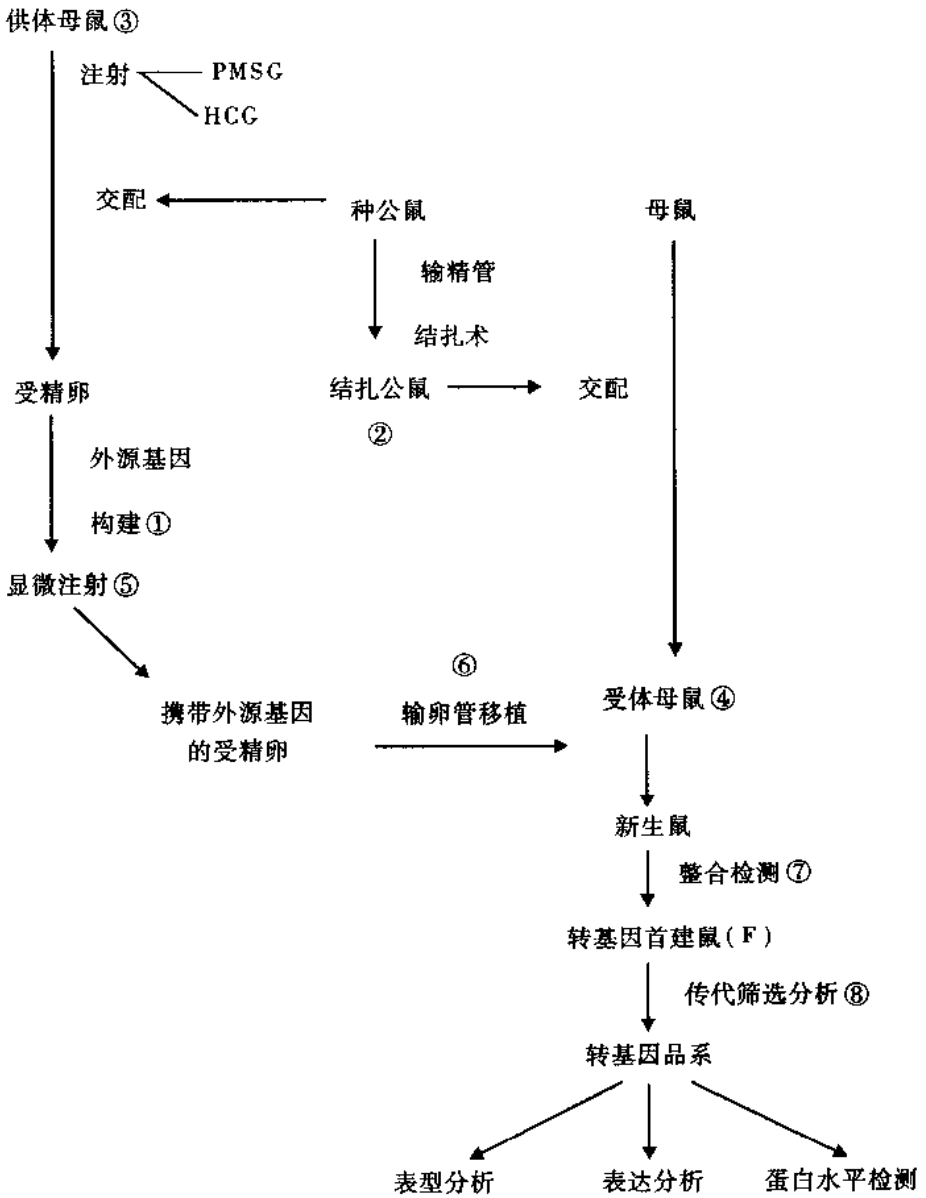


图 11－1　显微注射法产生转基因鼠的实验步骤

转基因效果的分析，主要可从下面3个方面进行衡量：

(1)整合检测：提取鼠尾DNA分别选用：①斑点杂交；②PCR分析；③PCR产物测序；④Southen blot。

(2)表达分析：提取组织RNA分别选用：①RT－PCR；②定量RT－PCR；③Northern blot；④定量RNA保护分析。

(3)蛋白水平检测：①酶活力；②放射分析；③Western blot；④免疫组化分析等。

外源DNA在整合位点主要是头尾相连排列。转基因的表达不仅与拷贝数有关，也与整合位点的染色体环境有关。近来，显微注射技术已用于下调转基因大、小鼠中的基因表达。这种方法先是构建一个从相反方向把一部分cDNA融合到启动子的目的基因，用显微注射方法产生转基因动物，使转基因表达反义RNA。在转基因动物细胞中两种RNA形成的双链将阻断mRNA的加工和翻译，从而阻止基因的表达。

显微注射法的主要优点：①在具备配套仪器和熟练操作技术之后，实现基因导入的速度快且操作简单；②由于基因是导入受精卵原核内，外源基因在宿主染色体上的整合率相对较高，较少产生嵌合体，有利于对当代转基因的表达分析，并能够快速地建立转基因动物品系；③导入外源基因片段的大小无限制，可长达250 kb等，且导入的基因也无需载体。不足之处：①外源基因的整合是随机的，故整合率不能检测，因整合引起的突变和位置效应(在转基因插入部位或附近存在有调节序列)可使转基因表达发生改变；②有的动物原核不易看清楚，如大鼠、猪、山羊等，需经特殊处理，才能有效导入。

(二)基因敲除技术

基因敲除技术(knockout technology)又称基因靶向技术或基因打靶(gene targeting)，也就是将构建的靶基因载体(简称靶载体，又称打靶载体)序列导入细胞内，这一序列就可以与该细胞基因组内的同源序列发生重组，靶载体被放置到细胞的基因组内，实现基因定点突变(即定点灭活)，建立起基因定点灭活的细胞系或基因定点灭活的动物。基因敲除技术自从八十年代末开始以来，已在诸多研究领域中得到应用。基因敲除的过程是先要在构建的靶载体的目的基因片段中插入一外源基因作为选择的标志基因并使目的基因灭活。将此靶载体导入既可培养又有发育全能的胚胎干细胞(embryonic stem cell，ESC，ES或Evans Kaufman cell，EK)中，通过与细胞内的基因同源重组，使灭活的基因取代染色体上原有的基因。利用标志基因产物的特殊药物抗性进行筛选和基因型鉴定，可以得到基因定点灭活的ES细胞。为使基因缺失的细胞能发育进入细胞系，将重组的ES细胞注入宿主胚胎胚泡腔内，植入假孕母体子宫里，发育成带有已定点灭活了内源目的基因的ES细胞的种系嵌合体。通过杂合的种系嵌合体的自交，再筛选出纯合的基因敲除动物。进一步对基因敲除动物的表型和基因型、基因表达及其产物进行分析，以研究被敲除基因在个体发育中的时空调节和基因缺陷所出现的表型异常(一些表型消失和一些表型出现)，建立疾病的动物模型。

靶基因载体的设计与ES细胞定点灭活基因的筛选是按下述原理进行的。在细胞内部，具有相似核苷酸顺序的DNA片段之间，可以相互配对并发生交换，称为DNA的同源重组(homologous recombination)。通常靶基因的同源序列在5～7 kb以上，同源序列一般应选择目的基因的功能区。同源序列中间插入的选择性标志基因，常用的有新霉素抗性基因(neomycin－resistance gene，neo)等。同源序列的一端或两端有时还可以连接上负筛选标志基因，如单纯疱疹病毒胸苷激酶基因(herpes simplex virus thymidine kinase，HSV－tk)。G418是新霉素的一种类似物，对真核和原核细胞均有毒性。细菌的*neo*能在真核细胞中表达。*neo*编码一种磷酸转移

酶，这种酶能使 G418 失活。当细胞表达这种 *neo* 基因后，就可在 G418 的选择培养基中存活。Gancyclovir(GANC)是一种核苷酸类似物，hsv－tk 表达产物可使 GANC 转变成一种有毒物质，使细胞死亡。因此，转入有 hsv－tk 的细胞不能在含有 GANC 的培养基中生存。当对同源重组后含有定点灭活的 ES 细胞进行筛选时，先要对在靶载体基因的一个外显子上插入的 neo 序列，用含抗生素 G418 的培养基培养筛选得到有 neo 的细胞，此为正选择(positive selection)。在同源重组过程中，由于 hsv－tk 是接在同源序列之外，不能整合到 ES 细胞上，而在随机插入的情况下，由于整个线性的外源 DNA 片段都可以插入到 ES 细胞基因组中，故 hsv－tk 能在转化后的细胞中保留下来。当细胞在含 GANC 的培养基中培养时，凡整合有 hsv－tk 的细胞就不能生存，而没有 hsv－tk 的细胞能得到富集，这个过程称为负选择(negative selection)。经此选择，外源基因未整合的或随机整合的细胞均不能存活，唯有定点整合的细胞存活。这就是目前广泛被采用的 G418 和 GANC 正负选择法(PNS)。目前应用得最多的是设计带有正、负筛选标志基因的靶载体。通常标志基因都带有能够在真核细胞中表达的启动子。建立基因缺陷动物的主要实验步骤如下：①构建靶基因载体；②分离培养 ES 细胞；③ES 细胞上的基因操作；用电转染法等将靶基因载体转染 ES 细胞；④获取胚泡期胚胎，以作 ES 细胞的移植受体；⑤用显微注射法将 ES 细胞克隆注入到胚泡腔内，与其内细胞团紧靠在一起，成为嵌合体；⑥将注射过的胚泡，经培养和筛选无发育缺损的囊胚，移植到交配第 3 d 的假孕受体鼠子宫内，培育出种系嵌合体动物，经传代培养得到纯合的基因缺陷的转基因动物。

基因敲除技术的主要优点：①可用多种方法将外源基因导入 ES 细胞，且鉴定与筛选比较方便；②采用基因打靶定点突变，故转入外源基因整合率的可控程度高于显微注射法；③ES 细胞注入囊胚以及囊胚移植到子宫在操作上比较简便。不足之处：①ES 细胞系的建立、保存和维持是一种高难技术，除小鼠 ES 细胞外，至今还未能得到其他动物的 ES 细胞，从而使基因敲除的技术应用受限制；②由于动物所有器官在整个个体发育中发生了基因突变，我们只能检测基因敲除动物某个主要基因在个体发育早期所起的作用，不能研究动物成年后基因的功能。许多情况下是纯合子活不到成年阶段。最近发展了一种条件敲除技术(conditional knockout techniques)，并已成功地运用于 Cre－loxp 系统，这是从细菌噬菌体 P_1 得到的一个重组系统，可以使靶基因失去活性，从而使得我们能实现在特定的组织或发育阶段敲除某个特定的基因；③由于整体动物具有很强的宽容性，单基因缺陷在整体中常可被代偿，一定程度上也增加了建立转基因动物模型的复杂性；④所得个体为嵌合体，筛选、培育动物所需的时间较长；⑤由于 ES 细胞的培养条件严格，对技术和经费的要求很高。

目前，显微注射法是较普遍采用的获得转基因动物的一种方法，本章将在以后的两节中，结合我们在转基因工作中的实践体会，以转基因大鼠为例，对显微注射法再作较详细的介绍。

第二节　显微注射法的操作步骤

一、实验条件的准备

(一)仪器和器材

1. 显微注射仪。由一架倒置显微镜和左、右各一显微操作臂组成。仪器配有照相和监视屏幕设备，并与拉针仪和解剖显微镜配套放置和使用(图 11－2)。倒置显微镜用于显微注射，配有微分相干相差装置(DIC)。由 10～15 倍目镜，6.3×，10× 和 32× 物镜组成。显微操作臂

由左、右两臂组成，装在显微镜载物台的两侧，左侧接持卵器，右侧接显微注射针，分别用以控制持卵器和显微注射针在显微镜视野内的操作和调节它们的三维空间。注射针刺入受精卵有手动和电动两种方式，后者可用电脑来控制操作程序。

2. 解剖显微镜。

3. 拉针仪。

4. 微注射针和移植器具。

(1)持卵管(holding pipet)的制作：双手执毛细玻璃管的两端，将管的中部置酒精灯外焰烧红至变软后迅速离开火焰，同时双手外伸，将烧软的部分拉细。用玻璃刀或细砂轮小心将细管切开，在 100× 显微镜下观察切口应平齐（如不平齐，应重新切割）。然后于酒精灯火焰底部的蓝焰边缘处将切口钝化处理(即将管口于蓝焰边缘处做短暂灼烧，然后于显微镜下检查管口形状，如此反复，直至满意)。如果实验室配有持卵管制作仪(如日本 Narishge 公司的 MF－9 型加热仪等)，则可在显微镜下直接监视加热灯丝对持卵管管口灼烧后的形状，及时调整二者之间的相对位置，更易得到满意的持卵管。理想的口控持卵管的持卵端外径 50～100 μm，内径 20～30 μm，而泵(如注射器)控制的持卵管内径以小于 15 μm 为宜，细端长约 1～3 cm。

(2)移卵管(transfer pipet)的制作：用于受精卵的移植，基本同持卵管的制作(但可使用较粗的玻璃管)，只稍加钝化，使末端圆滑即可，理想的移卵管内径应约为 100～150 μm。移卵端钝化处理后，将另一端连于一适当长度的硅胶管上，硅胶管末端可接一次性口控管，构成“口控管－硅胶管－移卵管”的操作系统。

(3)注射针(microinjection needle)的制作：按拉针仪(puller)使用说明操作。注射针的尖部开口应小于 1 μm。

(4)洗卵管(flushing pipet)的制作：基本同移卵管的制作。洗卵管的内径约为 150 μm 为宜。

5. 其他用品。培替氏培养皿(Petri culture dish)，毛细玻璃管(外径 1 mm)，酒精灯和玻璃刀(或小砂轮)。

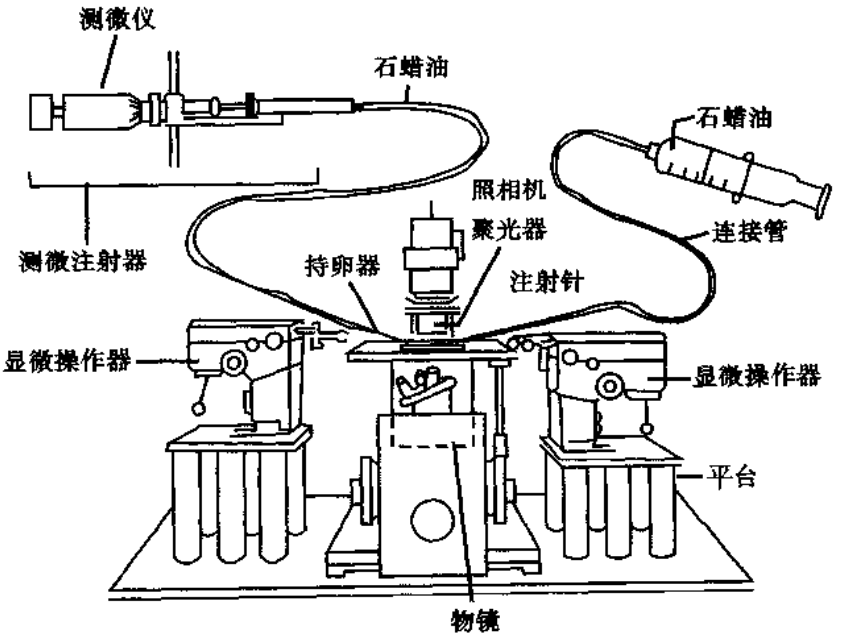


图 11－2　显微注射仪示意图

(二)试剂

1.培养液。培养液需要用重蒸去离子水配制,经微孔滤膜(0.22 μm)过滤除菌。滤器、滤膜应高压灭菌。目前实验室普遍采用两种培养液,即 M2 和 M16 培养液。

(1)M2 培养液:此种培养液(表 11－1),以 $NaHCO_3$、Hepes 增加缓冲能力,主要用于洗卵和受精卵注射用。

表 11－1　M2 培养液的组成及配制用量

成分	mmol/L	分子量	用量(g/L)
NaCl	94.66	58.450	5.533*
KCl	4.78	74.557	0.356
$CaCl_2 \cdot 2H_2O$	1.71	147.200	0.252
KH_2PO_4	1.19	136.091	0.162
$MgSO_4 \cdot 7H_2O$	1.19	246.500	0.293
$NaHCO_3$	4.15	84.020	0.349
Hepes	20.85	236.300	4.969
乳酸钠	23.28	112.100	2.610
			(或 60%糖浆 4.349g)
丙酮酸钠	0.33	110.000	0.036
葡萄糖	5.56	179.860	1.000
牛血清白蛋白			4.000
青霉素 G 钾盐(100 U/mL)			0.060
硫酸链霉素(50 μg/mL)			0.050
酚红			0.010
重蒸水			加至 1L

* 当需要无 Ca^{2+} 培养液而省去 Ca^{2+} 时,则 NaCl 增加至 5.68。

配制方法:

1)准确称取 hepes,溶解于 50～100 mL 重蒸水。

2)称取青霉素、硫酸链霉素及 $CaCl_2$,分别溶解于重蒸水。

3)称取除牛血清白蛋白和乳酸钠之外的其余成分,以约 500 mL 的重蒸水溶解后,置入 1 L 体积的容量瓶内。

4)将 1)、2)的溶液加到容量瓶中。

5)若是乳酸钠糖浆,则需用 10 mL 量筒准确量取后再加到容量瓶中,而后补以重蒸水使总体积达 1 L。

6)最后将牛血清白蛋白加在溶液界面上,使其慢慢溶解、混匀。注意不要振荡溶液,否则会影响蛋白质对有毒物质的吸附。

7)调 pH 至 7.2～7.4。

8)使溶液通过高压灭菌的滤器除菌。

9)分装于灭菌的瓶中,4 ℃保存 1 周,－20 ℃可保存较长时间。

(2)M16 培养液：主要用于受精卵在孵箱内培养时用。除不加 Hepes 外，其他成分同 M2。

2. 石蜡油(分析纯，不需灭菌处理)。

3. 透明质酸酶用 M2 培养液配成 10 mg/mL 的贮备液，过滤除菌，分装保存(－20 ℃)。

4. 孕马血清促性腺激素(PMSG)用灭菌生理盐水配成 100 IU/mL，分装、避光保存(－20 ℃)。

5. 人绒毛膜促性腺激素(HCG)以灭菌生理盐水配成 100 IU/mL，分装避光保存(－20 ℃)。

6. 麻醉剂。采用长春解放军农牧大学兽医研究所生产的 846 合剂。

(三)外源基因

用于微注射的外源基因，纯度要求很高，应不含有颗粒，对受精卵无毒。可用线性化，50kb 以内的 DNA 片段。长片段的基因易引起断裂，制备和操作时应注意。制备和配制基因及试剂需用高纯度水，缓冲液为 10 mmol/L Tris，pH7.4，0.1～0.25 mmol/L EDTA。所用器皿要洁净。用于转基因的 DNA 样品一般是特定的基因片段，需经琼脂糖电泳后分离制得。凡适宜于从琼脂糖中分离纯化 DNA 的方法，均可用于显微注射用 DNA 的制备。但从琼脂糖中分离 DNA 后，需经充分纯化才能满足转基因的需要。从琼脂糖中分离 DNA 的方法较多，可根据分离片段的大小来选择使用。用于转基因的 DNA 片段一般较大，选用透析袋法，V－型槽法和玻璃粉法均可获得满意效果。基因浓度和整合拷贝数间无一定关系，多数认为应在 1～5 μg/mL，注射量在 1～2 pl.(微微升)。

(四)结扎雄鼠的准备

实验前将 300～350 g 的雄性大鼠作结扎输精管的节育手术。于术后两周可用于和雌鼠交配，以获得假孕的受体母鼠。通常可使用一年左右。雄鼠结扎输精管的手术操作如下：体重在 300～350 g 的雄性大鼠(Wistar 种)，以腹腔注射 846 合剂(0.1 mL/100 g 体重)麻醉。腹部皮肤去毛，酒精消毒，沿腹中线皮肤作 1.5～2.0 cm 的切口。用镊子沿腹壁下夹住白色脂肪组织后向上拉出切口，睾丸、输精管也随之被提出腹腔。在睾丸下方有一白色较硬的细管即输精管。一手持镊夹住输精管，另一手持尖头镊子将输精管周围的结缔组织分开，用烧红的镊子夹住这段输精管，即被烧断，完成结扎手术。将睾丸送回腹腔。重复此步操作以结扎另一侧输精管。切口处放少量青、链霉素粉剂后缝合；碘酒、酒精消毒；置鼠于洁净、温暖处，待其苏醒。

二、供体鼠和受体鼠的同步准备

(一)供体鼠

选择体重 230～250 g 的动情间期的雌鼠(Wistar 种大鼠)，于第 1 天中午 12 时腹腔注射 20 U 孕马血清促性腺激素(pregnant mare's serum gonadotropin，PMSG)，第 3 天中午 12 时再注射 20U 人绒毛膜促性腺激素(human chorionic gonadotropin，HCG)，同时将雌鼠与 350 g 体重的雄性大鼠按 1:1 比例合笼，一般当晚交配。第 4 天清晨查雌鼠阴道，出现阴栓者为已交配的供体鼠。第 4 天晚 7～8 时将其处死，获取超排的受精卵(光周期 7:00 A.M.～7:00 P.M.；暗周期 7:00 P.M.～7:00 A.M.)。供体鼠的交配率一般在 75%～100%。

注意：供体鼠制备中，为获取含有极体的优质受精卵，需采用性成熟的正常雄鼠(约 350g 体重)作为种鼠与雌鼠交配。种鼠在两次交配间应间隔 3 天，还应严格控制光、暗周期。卵的超排数量与激素用量有关，需按体重计算。不同厂家及批号的产品，需作预实验后确定最佳用量。一般每只大鼠排卵 20～30 枚即可。将雌鼠放入雄鼠居住过的笼中，可提高交配率。

(二)受体鼠

在供体鼠注射 HCG 的同时,将 230～250 g 处于动情前期－动情期的雌鼠(Wistar 种大鼠)与结扎雄鼠(按 1:1 比例)交配,次日晨与供体鼠一起检查阴道,有阴栓者即为假孕受体母鼠,可用于受精卵的同步移植。受体鼠的自然交配率在 25%～50%。

注意:受体鼠准备中,将数只动情后期的鼠集中同笼喂养几天,有助于动情期鼠的挑选,提高交配率。雌、雄(结扎)鼠按 2:1 合笼也有助于提高交配效率。

三、受精卵的采集

断头处死供体鼠。消毒腹部皮肤。在腹中央横向剪开皮肤,用手将皮肤分别向头、尾方向撕开,直到腹部完全暴露。剪开肌肉层,在左、右肾的下方找到红色的卵巢,周围有白色脂肪组织包裹。分离输卵管附近的脂肪及膜,用虹膜剪将输卵管剪下,置于盛有 200 μL M2 培养液的凹盘内。解剖显微镜下,左手用一尖头镊固定输卵管,右手用另一尖头镊细心将膨大壶腹部撕破一小口,可见一团白色粘性的包含有受精卵的卵丘细胞缓缓涌出。小心取出输卵管部分,置入含透明质酸酶(1 mg/mL)的培养液中,消化 5～6 min,用洗卵管移出游离的受精卵,放入新鲜的 M2 中洗 3 次,在 200× 显微镜下筛选出符合要求的受精卵。每次将 30～40 枚受精卵移入注射槽(如凹玻片)里的 M2 培养液中,用于微注射。注射槽上的 M2 培养液体积为 30～50 μL,并以石蜡油覆盖,以免在注射期间挥发。其余的受精卵移入 M16 培养液,于 37 ℃孵箱中培养待用。

四、受精卵的显微注射

(一)注射前准备

1.检查无菌吸水纸、玻璃刀、酒精灯、火柴、消毒碘酒和酒精是否备齐。

2.手术器械需煮沸或高压灭菌。平皿和凹玻片应高温处理(80 ℃, 2 h)以灭菌和除去 DNase。超净台和微注射室事先用紫外灯消毒。

3.用前将 M2 培养液放 CO_2 孵箱,37 ℃平衡 1 h。

4.微注射针及移植器具(除洗卵管外)均临用前制作。

(二)显微注射操作

1.安装微注射针及持卵器。持卵器和注射针分别安装在显微注射仪的左、右操作臂上。在低倍镜下调节持卵器和注射针的位置,使其在视野中清晰地位于同一平面,并在一条直线上。

2.吸入外源基因。将注射针小心地插入到外源基因(DNA)溶液中,使其充满针的尖端。将注射针放回到右操作臂上,与含液体石蜡的注射器连通。

3.注射槽的准备和受精卵的滴入。用凹玻片(或培替氏培养皿)作微注射槽,在超净台内将 M2 培养液滴入凹玻片中央,液滴的直径不超过 1 cm,其液面要水平,以避免折射效应,在 M2 液滴上覆盖一层石蜡油,将其置于倒置显微镜的载物台上,于低倍镜(6.3×10)下,调节焦距至清楚为止。

用洗卵管吸出在 37 ℃孵箱中事先用 M16 培养的供体卵,在解剖镜下移入 M2 培养液中洗涤两次,放入凹玻片液滴上部。调焦距使在低倍镜下清楚地看到受精卵的轮廓,并用持卵器将卵汇聚到一起。

4.调节卵平面。重新调整焦距,让持卵器、卵、注射针位于同一平面内,以便自如地操作卵。移动左操作臂,让持卵器对准一枚卵,增加负压,使持卵器末端吸住受精卵,再移动持卵器到视野中央,转换高倍镜(32×10)或用40×DIC物镜,调节焦距到两个原核清晰为止。

5.注射外源基因。移动右操作臂,使注射针的针头紧靠受精卵的透明带,调节操作臂上下控制旋钮,使针与原核处于同一平面上,快速将针刺入受精卵雄原核内,旋动注射器,将基因注入原核内(图11-3)。若见原核膨胀,则继续注射,使体积达到原来的两倍。注射基因的量为1~2 pL,迅速撤出注射针。转换到低倍镜(3.2×10)下,移动持卵器,将已被注射基因的卵移至视野内在持卵器和注射针以外的另一区域,以区分未注射的卵。重复上述操作,注射另一枚卵。一根注射针一般可注射5~10枚卵。将注射后的受精卵移至M16培养液中,洗涤两次,培养至移入假孕母鼠前为止。

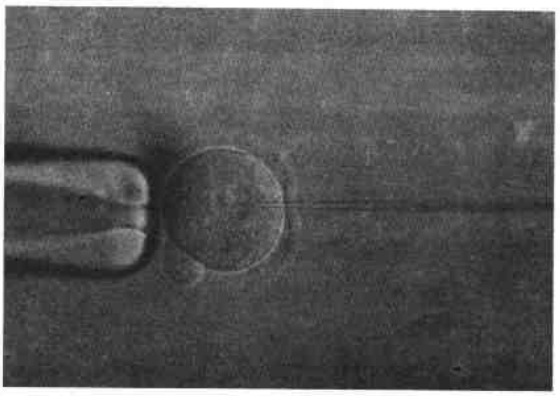


图11-3 小鼠受精卵的显微注射

(注:盛有外源基因的注射针正刺入小鼠受精卵的雄原核中)

注意:受精卵中有雄原核和雌原核。原核中可见核仁。通常雄原核较大并位于极体的对侧。微注射中正确分辨受精卵的雄原核,并将外源基因准确注射到雄原核中是能否获得转基因鼠的关键。此外,注射时微注射针不要触及核仁,避免反复刺伤透明带。

五、转基因后受精卵的再植

(一)受精卵的转移

小心地将注射有基因的卵吸入移卵管,卵的两端均有气泡和M2液间隔。

(二)暴露输卵管

将麻醉的受体鼠自然放置,在背侧最后一根肋骨后凹陷处(相当于靠近脊柱左侧或右侧1 cm处)(此处正对腹腔内卵巢位置),去毛消毒,作1.5 cm长的纵向切口,开腹后,用镊子沿腹壁将卵巢的白色脂肪夹住提出,即可看见红色卵巢、子宫和输卵管,将其取出置切口处。卵巢和输卵管不宜拉出太多,否则会影响视野且有可能产生损伤。

(三)找喇叭口(输卵管伞部)

解剖镜下,将卵巢和输卵管置于适当的操作部位。在血管稀疏处用眼科镊轻轻撕开卵巢和输卵管之间的囊膜,裂口不大于3 mm。吸净体液,左手用镊子伸入囊膜内,轻轻将卵巢和输卵管分开,可看见其中的输卵管伞部(多数受孕母鼠的伞部为白色、丰满的盲端)。右手镊子夹

住伞部顶端，上提至裂口处，交与左手镊，左手镊子轻轻夹住伞部边缘处，可见盲端中央有一小口。

（四）移卵

右手先将移卵管的口控端置于口中，再将移卵管从伞口轻轻插入伞部，在解剖镜下插入约 0.5 cm，小心移动移卵管，已确认移卵管已插入伞部。用口轻轻将受精卵吹入壶腹部（至第三个气泡至伞口）。轻轻拉出移卵管。将卵巢和输卵管放回腹腔。

（五）缝合

切口放少量青、链霉素，缝合切口。碘酒和酒精灭菌。将鼠编号，置于洁净处，待其苏醒。约 21 d 分娩。幼鼠出生 3～4 周后即可进行检测。

注意：

1. 移卵管的长度和内径要合适。选用 2～3 cm 长，90～100 μm 内径为宜。过长，吹、吸卵时阻力较大，不易控制；太粗，不易插入喇叭口，且卵易滑落丢失。

2. 植卵数：以 20 枚为宜。过少，因受精卵在微注射后有一定损伤，则成胎数少；过多，成胎率高（或过少所造成的个体大）引起难产或死胎。

3. 囊膜开口时应避免出血。开口应选择无血管区，小心撕开。如果血管过于丰富，应尽可能选血管较少处。在开口前或一旦撕裂血管后，均可分别夹住两边血管或该血管的上部约 2 min，以避免出血。

4. 提夹输卵管伞部时不可用力。用力过大，伞部断裂，移卵管难以插入；用力过小，伞部易脱落，脱落伞部易回缩，不易再找到。

5. 移卵管插入壶腹部时应轻、缓。移卵管口部要平齐、钝化，应试探性插入，不可用力，否则穿透壶腹部而造成移植失败。

第三节　转基因整合与表达的检测

一、基因组 DNA 的提取

根据所用检测方法的要求，尽可能提取纯化，以提高检测反应的灵敏度和效果。

以下简要介绍从鼠尾分离基因组 DNA 的步骤。

1. 切取 1 月龄子鼠尾尖 1 cm（用作 Southern 印迹杂交分析）或取 2～3 周龄子鼠尾尖 0.5 cm（用作 PCR 或斑点杂交分析）（通常 10～20 mg，可提取 10～20 μg DNA）用蒸馏水洗净、粗剪后，放入含 660 μL 水解液（50 mmol/LTris－HCl，pH 8.0，100 mmol/L EDTA，100 mmol/L NaCl 和 1%SDS）内。

2. 40 μL 蛋白酶 K（proteinase K）（10 mg/mL），50～55 ℃过夜，间或摇动。

3. 加入 700 μL Tris 饱和酚，轻轻混匀（约 20 min），使蛋白质变性。离心（在微型离心机上，10 000 r/min，10 min），置上层黏稠的水相到一新管（或用 Tris－饱和酚同法再抽提 1 次）。

4. 用约 600 μL（等体积）酚/氯仿/异戊醇和约 600 μL 氯仿/异戊醇各抽提 1 次。

5. 加 0.6 倍体积异丙醇，轻轻混合（可反复颠倒试管约 30 s），以加样用塑料滴头（tip）挑取白色、絮状的沉淀。

6. 以 1 mL 70%冷乙醇漂洗沉淀 1～2 次，再将絮状沉淀移至另一新的微型离心管中，自然或真空干燥（注意不宜过干，否则难以溶解）后悬浮于 100 μL TE 中（可置于 55 ℃水浴中，1 h 即

溶解)。

7. DNA 样品的电泳检测和初步定量。

8. 注意:本方法操作简便,基因组 DNA 完整,得率高,但未除 RNA。通常 RNA 不影响杂交和 PCR 实验,故可不除去。需要时按下列方法除去 RNA。

9. 将上述 DNA 悬浮于 200 μL TE 中,加 2 μL 无 DNase 的 RNase(10 mg/mL),37 ℃消化 3~4 h。

10. 补加 300 μL TE,用酚/氯仿各抽提 1 次。

11. 加入 50 μL 3 mol/L NaAC,pH 6.0 和 2 倍体积乙醇,室温混匀,可见 DNA 絮状沉淀。

12. 取 tip 尖,小心将 DNA 挑入 70%乙醇中。

13. 按常规方法沉淀、洗涤 DNA,最后再悬浮于 100 μL 的 TE 中。

该法可除去大部分 RNA,需要进一步纯化时,可用 CsCl 密度梯度离心除之。

二、外源基因的检测

随着分子生物学方法的迅速发展,新方法不断应用到转基因检测中,这里仅扼要介绍常用的检测方法。

(一)染色体和基因水平

检测外源基因是否整合于动物的基因组以及整合的位点和拷贝数。

1.外源基因整合的检测。

(1)斑点杂交(dot blotting)。当目的基因与内源基因组 DNA 无同源性时可选用。本法对样品纯度要求低,快速、简便、经济,尤其对于大批子代动物的粗筛时颇具优越性,可作为首选方法。

(2)多聚酶链反应(PCR)。这是近年来常用的一种快速、灵敏的检测方法,只需极微量 DNA。但本法用于转基因检测时,要求引物特异性高。在引物设计方面我们的经验是:①可设计两对 PCR 引物,其中之一扩增内源基因片段,另一对则用以扩增基因的特定片段,且内源性 PCR 产物的长度应大于转基因的 PCR 产物的长度。这样,内源性 PCR 可以做为本反应管 PCR 成功与否的标志,而扩增转基因片段的 PCR 做为检测阳性动物的标志。②转入的基因是融合基因,则引物可分别设计在来自不同基因的片段中,这样特异性更高。③转基因的内源性基因的编码区同源性很高时,应注意其内含子长度是否相同,如果不同则可将引物设计在该内含子的两端,在 PCR 检测时由于其转基因和内源基因扩增的产物长度不同,可以判断是否有转基因存在。④根据 3′端不配对则不能延伸的原则,设计引物时应注意 3′端至少要有 3 个以上碱基不配对,而引物内应保证有最大数量的碱基不配对,以避免假阳性 PCR 产物产生。⑤还可根据转基因与内源性基因 PCR 产物是否具有酶切片段多态性,以区分之。此外,选用 0.1~0.2 μg 的基因组 DNA 已足够 PCR 扩增,用量过大,则杂质浓度相应增加,非特异性扩增也增强。PCR 变性中变性温度比时间因素更重要,宜采用"高温短时变性"(例如根据情况可选用 94 ℃,30 s 或 97 ℃,15 s 等)。

(3)PCR 产物测序法。显微注射产生转基因小鼠的过程中,外源基因是随机整合于小鼠基因组中的,整合过程中是否存在转入基因的丢失无法得知,为此我们建立了 PCR 产物测序的方法。与(2)中相同的上游引物经过同位素标记后进行 PCR 扩增。方法(2)中得到的 PCR 片段经低熔点胶回收后作为 PCR 扩增的模板。每个样品的一套 PCR 测序反应需 4 个反应管,分别为 T、C、G、A,加入相应的 d/ddNTP 混合物 2 μL。在另一离心管中混合模板 DNA 0.5~5 ng,

5×测序缓冲液 5 μL，标记引物 3 μL，以去离子水定容至 16 μL，再加 1 μL (5U/ μL)测序级 Taq DNA 聚合酶，混合均匀。分别取该模板混合物 4 μL 依次加到 d/ddNTP 混合物的管内，混匀后进行 PCR。PCR 程序为 95 ℃预变性 2 min，30 个循环为 95 ℃，30 s，51 ℃，30 s，72 ℃，1 min，最后 72 ℃延伸 10 min。每管加终止液 3 μL，80 ℃变性 2 min，冰浴中骤冷，进行 8%聚丙烯酰胺凝胶电泳，放射自显影后读取序列。本法可以进一步确定转基因阳性鼠中整合的外源基因的序列是否正确。本法具有操作简便，灵敏度高及特异性强等优点。

(4)Southern 印迹杂交。转入基因与内源性基因组 DNA 有较高同源性时，仍可选此法。本法准确性高。现将我们实验室有关转基因 Southern 印迹杂交洗膜条件的选择经验(表 11－2)介绍如下。

表 11－2　转基因 Southern 印迹杂交洗膜条件

洗膜液	洗膜温度(℃)				洗膜次数×时间(min)
	转基因同源性				
	100%	95%	90%	85%	
2×SSC	室温	室温	室温	室温	2×10
0.5×SSC,0.5%SDS	65	60	60	55	2×20
0.2×SSC,0.5%SDS			60	55	3×10
0.1×SSC,0.5%SDS	65	60			3×10
0.1×SSC	室温	室温	室温	室温	1×10

2.整合位点的检测——染色体原位杂交。用标记探针与培养转基因动物细胞的中期染色体 DNA 直接自显影后，即可确定转基因整合位点。

3.整合拷贝数的检测。可用斑点杂交(或 Southern 印迹杂交)，但和普通检测所用的斑点杂交不同，检测整合拷贝数的杂交是一种定量杂交，可计算出每个细胞中外源基因的拷贝数。

(二)转录水平

在 mRNA 水平检测外源基因是否表达。通常在作完整合检测，并有足够的子代动物数时进行。常用方法如下：若转基因与内源性基因同源性较小，可选 Northern 印迹杂交；若两者同源性较大，则选 RT－PCR 或 RNase 保护分析。

1. Northern 印迹杂交。当外源基因和转基因动物基因组 DNA 同源性较高时，Northern 杂交是确定外源基因是否被转录的有效方法。

2. 逆转录—聚合酶链反应(RT－PCR)。本法灵敏度高，对内、外源基因的同源性要求低，故比 Northern 法有较大适用范围，还可用内标准进行定量。

3. RNase 保护分析法。需要体外转录合成一段与待测 mRNA 顺序互补的单链的 RNA 探针(高比放射活性)方可检测。和 Northern 法相比，本法更为准确，故被广泛应用于基因转录水平的检测。

(三)翻译水平

在蛋白质水平检测转基因是否表达。它包括两个方面，即转基因的 mRNA 是否被翻译和翻译的蛋白是否具有生物功能。常用方法如下：

1. 免疫检测技术。用免疫学方法检测蛋白质类表达产物的存在，如 Western 免疫印迹（见参考文献 12），ELISA 分析（免疫酶联测定法），放射免疫分析法和免疫荧光法等。

2. 生化性质和生物学活性的分析。除直接测定基因表达产物外，还可通过测定表达产物的理化性质和生物学活性来鉴定表达产物的存在。如酶活力测定，受体蛋白分析等。

3. 转基因动物整体水平观察。从遗传学上，在动物的整体水平观察基因型和表型的关系以及分析转基因对动物整体性状和生理功能的影响。

总之，一个理想的检测方法的选定始于构建外源基因时，即要考虑到将来检测的方便尽可能设计、构建包含与内源基因结构不同的部分，以增加非同源性。另外，要考虑检测样品的数量和所需的费用。若被检测基因与内源基因的同源性较小，数量大，首先在整合检测方面，可先选斑点杂交进行粗筛，然后作 Southern 印迹杂交，对完整的基因进行鉴定，以确定外源基因是否整合，并用染色体原位杂交和定量斑点杂交分别作整合位点和拷贝数的鉴定；其次，可选 Northern 印迹杂交作转录水平的检测；最后根据所表达蛋白的种类和性质以及功能，进行蛋白水平的检测。若被检测的基因与内源基因的同源性较大，则在基因和转录水平的检测方面所选的方法与上述不同，可选 Southern 印迹杂交（或选用适当的限制性内切酶进行酶切后再作 Southern 杂交分析，例如，我们曾将小鼠的肾素 - 2 基因转入大鼠基因组中，两者有 85% 的同源性，但当用 *EcoR* Ⅰ进行酶切时，则肾素 - 2 基因可产生 9.2 和 4.4 kb 的两个片段，而大鼠的肾素基因则产生 6.5、4.4、2.4、1.1 和 0.45 kb 的杂交带，显然用 *EcoR* Ⅰ酶切后所得 9.2 kb 片段为转基因的特征片段）或选设计有高特异性引物的 PCR 法作外源基因是否整合的鉴定，选用 RT - PCR 法或 RNase 保护分析法作转录表达的检测。

体会和展望：

1. 开展转基因动物工作的主要体会：

（1）显微注射法制备转基因动物技术，从方法学角度考虑，可称得上是一种“系统工程”。它包括从动物实验技术，显微注射技术和相应于这些实验条件的各种准备方法以及分子生物学方法。每一部分又包括若干技术，各部分实验必需同步设计和配套。为此，要求从事该项工作的人员具有良好的科学和心理素质，在工作中应有高度的责任心，周密的计划，熟练的技巧和细心的操作，各个环节都不得失误，否则将达不到预期的目的。

（2）转基因个体的检测和判断也很重要，要求所选用的检测方法必需准确、灵敏，否则因判断错误也会造成重大的失误。

2. 目前转基因动物技术本身存在的问题主要有以下几个方面：

（1）转基因的阳性率低。通常转基因首建的阳性率在 10% ~ 20%，有时甚至低于 1%。

（2）转基因的整合造成宿主细胞基因的突变。有插入突变，缺失突变，宿主基因扩增和位移。

（3）转基因的表达不能预测。主要表现在转基因的表达紊乱，包括不在预期组织中表达；同一启动子在不同的个体或不同的动物中表达行为不同。

（4）转基因动物模型可能和预期不符。

3. 未来转基因动物研究可能集中在以下几个方面：

（1）转基因动物技术的不断完善。如改进现有的显微操作技术和提高外源基因整合率；广泛开展用基因敲除技术并建立转基因小鼠和探索大鼠的胚胎干细胞培养方法和建立基因敲除的转基因大鼠模型的方法。

（2）深入开展转基因的表达调控研究。从转单基因的动物的组织特异性表达调控，发育

阶段的时空调控到建立携带多个基因的转基因动物模型，以研究基因间相互作用和多级调控。建立拼接特定的组织特异性启动子的单、双基因的转基因动物。

（3）建立用于药物研究和适合基因治疗的转基因动物模型也将成为今后发展的重要方向。

（陈兰英）

第十二章

RNA 干扰

内容提要

本章简述了 RNA 干扰的原理、实验操作及其在生物医学和基因组学中的应用。RNA 干扰是 dsRNA 诱导的特定 mRNA 的降解过程，是通过在转录后水平降解靶 mRNAs 而实现的。其自然功能是保护基因组防止像转座子和病毒那样的可转移遗传元件的入侵。RNA 干扰在基因功能研究、信号传导通路、遗传学筛选、转基因动物、基因治疗等生物医学和基因组学研究和应用领域具有十分广阔的前景。

研究基因的功能可分两种策略：即表型驱动法和基因驱动法。前者包括 ENU(ethylnitrosourea)大规模点诱变和基因捕获。基因驱动法包括目前很流行的基因敲除(knock out)/敲入(knock in)等。而近年来广为应用并有可能引发哺乳类体细胞遗传学革命的方法是外源双链 RNA(dsRNA)依赖的转录后基因沉寂或 RNA 干扰(RNA interference，RNAi)，它极大地丰富了功能基因组的研究手段和内容。因其在转录后水平而非基因组水平将特定基因 mRNA 的表达下调，故有人将其称为敲低(knock down)。

RNA 干扰首先是在秀丽线虫中作为 dsRNA 的生物学反应 - 诱导序列特异的基因沉寂而被发现的。RNA 干扰或 RNA 沉寂是一双链 RNA(dsRNA)诱导的同源性依赖的特定 mRNA 降解过程。RNAi 的一个关键特征是其通过在转录后水平降解靶 mRNA 实现其效应。

第一节　RNA 干扰的原理

遗传和生化研究表明，dsRNA 诱导的沉寂现象出现在各种生物，包括植物、真菌、原生动物等，而且具有共同机制。当前的看法是：RNAi 是由作为 RNAi 引发者的 dsRNA 被核酸内切酶 Dicer 降解为 siRNA(或称引导 RNA，gRNA)开始，且依种系不同，其长度范围在 21 ~ 25 nt 之间。这些 RNA 整合入多组分核酸复合物中，该核酸复合物又利用 siRNA 的反义链序列识别并降解同源 mRNA。其具体机制如下：

转染的或胞内产生的、作为 RNAi 引发者的 dsRNA 首先由 RNase III 核酸内切酶 Dicer 切成大小在 21 nt 左右的双链 siRNA，切除反应和 siRNA 的产生需要 ATP，这些 siRNA 的正义和反义链的 3′端都有 2 nt 的突出悬挂碱基。与靶 mRNA 互补的 siRNA 链的 5′磷酸基团是随后的 RNAi 所必需的，而与靶 mRNA 序列相同的 siRNA 链的 5′磷酸基团则为 RNAi 部分需要。自由

的 3′羟基是随后的 RdRP(RNA - dependent RNA polymerase)反应所必需的。siRNA 有 3 个特征：①21 ~ 23 nt 长度；②2 nt 3′突出端的双链结构；③5′磷酸基团。这些特征允许细胞区分真正的 siRNA 和"冒名顶替"的 siRNA，此时 siRNA 特异的激酶将维系进入 RNAi 通路的、真正的 siRNA 的 5′磷酸化状态，而不将 5′磷酸基团加于其他小的 RNA，以保证真正 siRNA 靶向 mRNA 并使之被降解。在达到阈值后，siRNA 的反义链与靶 mRNA 的序列以同源互补结合，并引导 RdRP 反应。由 RdRP 反应新合成的 dsRNA 再由核酸内切酶 Dicer 酶切并进一步 siRNA 扩增。一部分新合成的 dsRNA 可能由 RNA 解旋酶或 RdRP 本身(因为 RdRP 含 RNA 解旋酶结构域，所以可能有解旋酶活性。)解旋，并以其正义和反义链为模板，再由 siRNA 作引物由 RdRP 经历下一轮的扩增。

这就起始了 siRNA 和模板 RNA 分子的指数扩增过程，因此少量 dsRNA 即足以在很长时间灭活连续转录的靶 mRNA 分子。在植物中，这种灭活效应可扩散至未处理的细胞和组织。如在黑腹果蝇中，siRNA 可被整合入一 360 kDa 的复合物，该复合物不具靶向切除活性。在提供 ATP 的情况下，由于 RNA 解旋酶的作用，360 kDa 复合物变为具有活性的多亚单位复合物(RNA - induced silencing complex，RISC)，介导随后的单链 mRNA 降解反应。其中 siRNA 的作用是通过与靶 mRNA 序列同源的序列碱基配对，从而识别靶 mRNA，由 siRNA 分子引导 RISC 复合物中具有 RNase 活性的部分(与上述核酸酶 Dicer 不同)，以降解与 dsRNA 相关的 mRNA。

第二节　RNA 干扰的生物功能

RANi 的一个自然功能是保护基因组免受内源性转座元件破坏。细胞中转座子以 10 至成千上万的拷贝数出现，当转座子以反义方向整合入内源性启动子时，产生转座子 dsRNA。与转座子相似，转基因整合入基因组是一相对随机的事件，可激活与引入与转基因序列相似的 PTGS(post transcriptional gene silencing)。植物中胞浆复制的 RNA 病毒亦充当 PTGS 的靶子和诱导物，因此提示 PTGS 的另一功能即抗病毒防御。与这一假说一致的是，发现某些植物病毒编码的蛋白能抑制 PTGS。

RNAi 的自然功能和共抑制被认为是保护基因组，防止像转座子和病毒那样的可移动遗传元件入侵。激活的转座子和病毒可在宿主细胞内产生不正常 RNA 或 dsRNA，因此特定 mRNA 的降解可以防止转座子和病毒的复制。

第三节　哺乳类中的 RNA 干扰

一、RNAi 在哺乳动物中存在的证据

RNAi 可能是基因沉寂的通用机制，因为至今除了酿酒酵母外，在其他所有的基因组中都发现有参与沉寂过程的同源蛋白，提示哺乳类也应存在 RNAi。

RNAi 存在于哺乳类动物的最早证据是 dsRNA 可诱导小鼠卵母细胞和早期胚胎序列特异的沉寂。然而用长度在 38 ~ 1 662 bp 的 dsRNA 却一直未能在常用哺乳类动物细胞系，包括 293、NIH/3T3 等中引发序列特异的 RNAi，这是始料未及的。

二、哺乳动物中的 RNAi

早已知道，哺乳动物有强大的抗病毒反应系统，哺乳动物的胞浆 dsRNA 会激活哺乳动物

细胞非特异的病毒防御机制－干扰素反应。当细胞感染病毒或给予 dsRNA 时，细胞分泌干扰素，干扰素反应是序列非特异的，通常被长度大于 80 bp 的 dsRNA 激活，其效应主要通过 PKR 和 2′－5′寡腺苷聚合酶完成。在有 dsRNA 时，PKR 通过二聚化激活。随后 PKR 磷酸化 EIF2α，引起一非特异的翻译关闭。同时 dsRNA 也激活 2′－5′寡腺苷聚合酶，该酶产物是非特异核酸酶 RNase L 的必需辅因子，RNase L 可以降解 mRNA。这就是序列非特异的 dsRNA 诱导的机制。干扰素反应将导致 RNA 转录本的非特异降解和宿主蛋白翻译的关闭。由于这一障碍，RNAi 似乎不是哺乳动物细胞中沉寂基因的通用方法。

但最近 Tuschl 等人显示可通过引入 siRNA，在无数哺乳动物细胞株中激活了 RNAi，而且他们发现可以用体外合成的、长度为 21 nt 的 siRNA 在哺乳动物细胞中进行基因特异的抑制，但由于序列短，可逃避宿主干扰素反应。这种体外合成的、有活性的 siRNA，是长 21 nt 的 dsRNA，并在 3′有对称的 2～3 个悬挂碱基的 siRNA。通过模拟 Dicer 酶的产物（Dicer 催化 RNAi 的起始步骤）而直接参与 RNAi 通路。而当 dsRNA 序列长于 30 nt 时，则可诱发干扰素反应。

三、长 dsRNA 引发的 RNAi

很早已知，哺乳动物发育早期，存在由 dsRNA 诱导的非特异反应减弱现象。研究发现在体外培养的未分化小鼠细胞中，长的 dsRNA 可诱导 RNAi 反应，包括小鼠 P19 和 F9 细胞。在小鼠胚胎干细胞中抑制靶 mRNA 可达 70%，但在已分化的细胞中没有该现象。这可能是因为与小鼠卵母细胞和早期胚胎一样，未分化的胚胎干细胞缺乏干扰素反应。这些实验结果说明长 dsRNA 的引发的 RNAi 存在于小鼠发育早期的细胞中。

多种小鼠胚胎源的细胞，包括正常胚胎干细胞也可激发 RNAi，而这些细胞通常用作遗传杂合（基因敲除/敲入）动物的前体细胞，这提示最终可以在整体动物水平验证基因沉寂效应。

四、哺乳类稳定 RNAi

仅仅用瞬时转染 dsRNAi 的方式在培养的哺乳类细胞或胚胎中实现序列特异的基因沉寂是瞬时的，由外源 siRNA 抑制蛋白合成也是瞬时的。经 siRNA 处理的细胞，靶基因的蛋白水平常在 siRNA 转染后 5～7 d 恢复正常（即 7～10 个细胞周期）。然而只有在能创造稳定的突变体时，才能实现最大的遗传学应用。长 dsRNA 诱导的稳定沉寂已在黑腹果蝇、植物、秀丽线虫中通过强制表达发夹 RNA 实现。

研究发现，短的、发夹 RNA（shRNA）可诱导哺乳类序列特异的沉寂。内源性表达的小的、瞬时的 RNA（stRNA）通过 RNAi 机制引发基因沉寂，以调节蛋白编码基因的表达。这些 stRNA 以短的发夹前体（约 70 nt）形式转录，并被 Dicer 处理为 21 nt 的 siRNA。siRNA 再通过碱基配对作用识别靶 mRNA。最近显示 shRNA 可在体外合成或在 RNA 聚合酶 III 启动子体内转录（下述），通过在细胞株中稳定筛选，得到稳定表达的发夹式 dsRNA 稳定的、序列特异的沉寂。这就为将来 RNAi 稳定遗传的转基因动物打下了坚定基础。

目前已在 MCF－7 和小鼠胚胎瘤 P19 细胞等细胞株中实现稳定表达。

第四节　RNA 干扰的实验过程

一、RNAi 的产生

目前发现，只有针对外显子序列的 siRNA 才能引发 RNAi。所以设计 RNAi 实验时，必须注

意这一点。一般从靶 mRNA 读码框中选 AA(N19)UU 型序列(N 为任一核苷酸),以获得对称的带 2 nt 的 3′突出悬挂一致序列。可以将 3′突出悬挂的 U 换为 2′－脱氧胸腺嘧啶以减少 RNA 合成的花费并增强对核酶的抗性。对 3′突出悬挂碱基,也可选用别的碱基,或者从 AA(N21)型序列中选择 siRNA,并将正义链的最后两个核苷酸换为 TT(因为据认为,siRNA 正义链的 2 nt 悬挂突出端可能不参与靶向识别)。在设计好序列之后,尚须进行 BLAST 相似性分析,以确证其序列的特异性。

目前而言,siRNA 产生途径有三种。

(一)体外转录

体外转录序列一般为 GN17CN2。含 T7 和/或 SP6 启动子,采用标准的体外 RNA 转录过程。T7 RNA 聚合酶的有效启动需 RNA 的第一个碱基为 G,转录后最后 2 nt 形成 siRNA 二聚体 3′的悬挂碱基,在正义链可变为 U。分别合成单链后,退火即成 siRNA。体外转录远比 RNA 合成经济。

(二)化学合成

化学合成 siRNA 的代价较昂贵。可以用 RNA 合成仪(如 Model 394, Applied Biosystem, Foster City CA),按厂家操作说明将 RNA 去保护后(例如可在以下条件去保护:3 mL 的 32%的氨/乙醇(3:1),于 55 ℃ 数小时),将 RNA 用 15%变性丙烯酰胺胶纯化,并从胶中洗脱,用 NAP－10 柱(Amersham Pharcia)去盐,真空干燥。dsRNA 退火一般采用酚/氯仿抽提的方法:将等摩尔浓度的正义和反义 RNA(依 dsRNA 长度不同,浓度在 50 nm 到 10 μm 之间)于 0.3 M NaOAc (pH6)中,90 ℃ 30 s,随后在室温下用等体积酚/氯仿抽提,再以氯仿抽提以除去残存酚。所得 dsRNA 用 2.5～3 倍体积的无水乙醇沉淀,团块在裂解缓冲液中溶解(100 mM KCl,30 mM HEPES－KOH,pH7.4,2 mM $Mg(OAc)_2$),dsRNA 质量可用标准的琼脂糖电泳检测(1 × TAE)。带 17 nt和 20 nt 3′悬挂突出端的 52 bp 的 dsRNA 于 95 ℃ 1 min,非常快速地降至 70 ℃,在 3 h 内缓慢降至室温(50 μL 退火反应体系,1 μM 链浓度,300 mM NaCl,10 mM Tris－HCl pH7.5)。退火后的 dsRNA 再次用酚/氯仿抽提,乙醇沉淀,并溶于裂解缓冲液中。

这里再提供另外一退火条件供参考:悬于退火缓冲液(100 mM KAc,30 mM HEPES－KOH, pH7.4, 2 mM $MgAc_2$),90 ℃ 1 min, 37 ℃ 1 h。对 50～500 bp 的 dsRNA 则于 37 ℃过夜,50 bp dsRNAs 的浓度为 8.4 μM,500 bp dsRNAs 的浓度为 0.84 μM。退火缓冲液如下:PBS, pH6.8, 2 mM $MgCl_2$,退火条件:RNA 寡聚物在 95 ℃孵育 1 min,降温至 70 ℃并慢慢冷至 4 ℃中,放置 2 h(10 μM)。

(三)体内表达(即由转入体内的质粒产生)

这是一种最近才出现的方法。

直至最近,妨碍 RNAi 广泛应用的原因是由于转入细胞的 siRNA 的瞬时性和 siRNA 需化学或酶体外合成获得。胞内转录 siRNA 的 DNA 模板带有 U6 或 H1 RNA 启动子,U6 或 H1 RNA 启动子都属于 RNA 聚合酶 III 启动子,二者大小不同,但都含相同的保守序列或反式作用因子结合位点,U6 启动子的 +1 位一定是 G(所以起始碱基最后为 GG),H1 是 A。当聚合酶 III 遇到连续 4 或 5 个 T 时(5 个连续的 T 作为 RNA 聚合酶 III 的终止信号),RNA 转录终止。

用质粒体内表达 siRNA 有两种方法(图 12－1):第一种是 siRNA 的正义和反义链分别由两个启动子转录。第二种是以发夹 dsRNA 的形式表达,经处理后产生 siRNA。其中发夹的大小、方向和序列对 RNAi 的效率很重要,中间发夹序列是由非同源序列构成。尽管核表达的 RNAs

活性形式还需长期研究，但就目前结果而言，推荐 siRNA 转录的模板：首先是正义链，随后是 4 个碱基（例 UUCG 发夹环），再后是反义链，并以 3′悬挂 Poly(U)终止子结束。然而有人发现，当发夹环为 5 个碱基时无效，6 个碱基时成功，7 个碱基时中等程度有效，9 个碱基的发夹环可有效引发 RNAi。看来其中的机制还远未阐明。

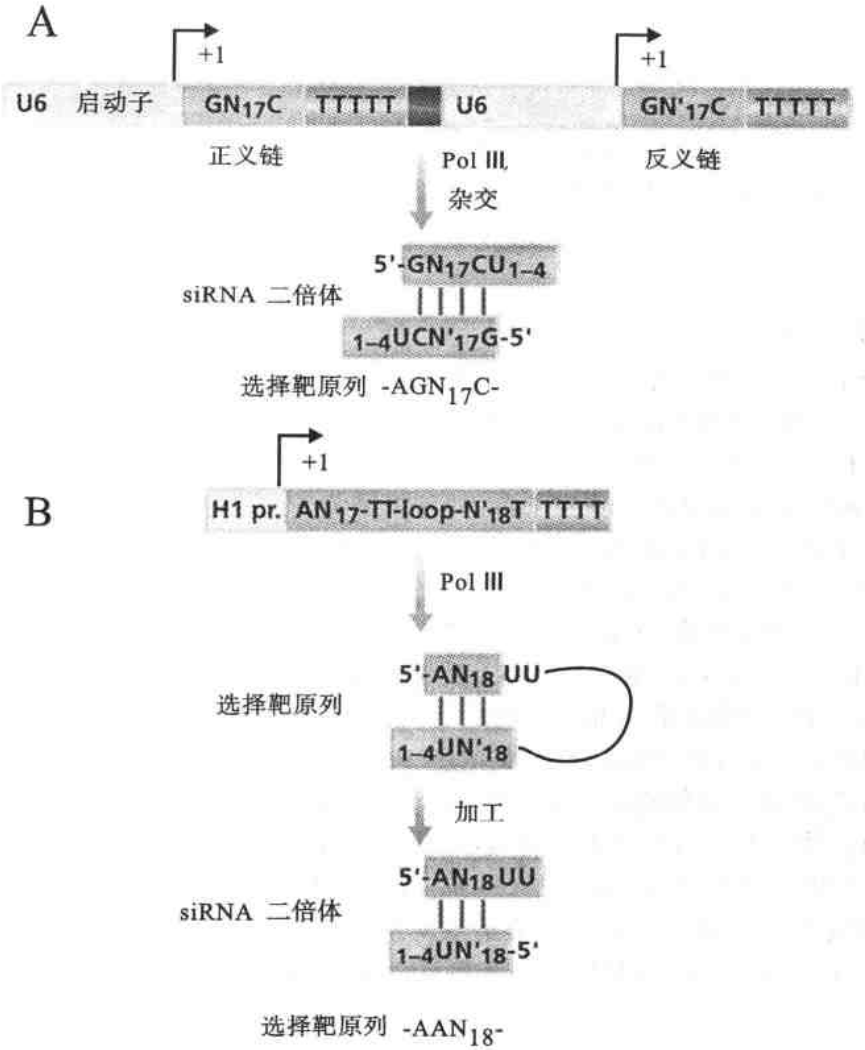


图 12－1　质粒体内表达 siRNA 的两种策略

（摘自 Tuschl T. Expanding small RNA interference. Nat Biotechnol. 2002 20(5):446～448.）

正如前面所述，siRNA 介导的靶向降解，需将其一或两条链的 5′末端磷酸化处理，这可能对其有效核酸蛋白复合物所必需。然而有结果显示，对发夹 siRNA 而言，也许不用将任何一链的 5′末端磷酸化处理，这些含发夹 siRNA 的 U6 转录本就能组装成任何必需的蛋白复合物。

基于载体的方法在多种哺乳动物细胞系中如 HeLa、H1299 和 C－33A 中成功应用，由质粒 DNA 表达产生的内源性 siRNA 克服了外源性 siRNA 给予途径的不足，尤其是其表型的易逝性。

若从稳定表达考虑，两个启动子表达 siRNA 的表达盒整合入基因组中，较产生发夹型 dsRNA 的表达盒稳定（因其含一不稳定的回文序列）。

用质粒而非合成的 siRNA 的优点：无 RNase 污染，比较便宜。但也有不利的一面，构建载体比较费时；与 siRNA 相比，效率低；而且在给予 siRNA 后 1～3 d，靶向必需基因引致的细胞生长抑制或死亡立即出现，因此在许多情况下不必要长期表达 siRNA。

二、给予途径

总体来讲，给予途径有三种，第一种是最常见的试剂转染法。转染 siRNA 用 Oligofectamine（专门用于转染短的寡核苷酸），也有用磷酸钙转染的。转染一般的 dsRNA 或质粒可以用 Lipofectamine 2000 等常规转染试剂。向 HeLa、小鼠 3T3 细胞和大鼠成纤维细胞中转染比较容易。

若细胞的转染效率不高，例如转染胚胎干细胞时，需要用电转或显微注射的方法。

若用稳定筛选，则可以根据载体抗性用 G418 等筛选单克隆，时间在 1～2 周。

三、沉寂效应鉴定

体内产生的正义和反义 RNA 用 Northern Blot 检测，一般是在转染后 24 h 提取总 RNA。由于 RNA 片段都比较小，所以一般用 18% 凝胶电泳，转膜。杂交条件也与常规 Northern Blot 检测不同，温度低，一般用 36 ℃。

若用原位杂交法检测，则在转染后 2 d 进行。检测蛋白表达一般在转染后 3 d。

基因沉寂后新表型可因具体基因不同而有所不同。

四、存在问题

目前发现，RNAi 存在如下问题：

1. 由于 RNAi 尚有许多未明的机制，有时必须沿 mRNA 试验几个靶位点以发现能产生最大抑制效应的靶序列，而且其中编码序列常由经验决定，起于 GG。

2. 由于 RNAi 是在转录后水平起作用，而对已翻译的蛋白不起作用，所以对那些半衰期长的蛋白来说，效果不是特别好。

3. 由于基于 siRNA 的技术是敲低（knock down）而非敲除（knock out）靶基因，如果靶基因是那些具有酶活性的蛋白而非结构蛋白，那么只要残留一点靶基因，RNAi 的表型难以鉴定。

4. 另外不同的基因似乎对 RNAi 的敏感性不同。例如在果蝇中表达于生殖系统的基因尤其对 RNAi 敏感，这种反应性表型可在后续几代中观察到，而对其他基因的效应却没那么持久。而有些基因难以用 RNAi 沉寂，如那些表达于产生成年结构组织（如翅膀、腿、眼、脑）的基因。通过 dsRNA 显微注射很难沉寂，仅以发夹环 RNA 形式表达产生 dsRNA 才行。不管哪种情况，野生型基因活性最终会被恢复，这可能是由于 siRNA 在细胞分裂过程中的稀释和降解。

第五节　RNA 干扰在生物医学和功能基因组中的应用

由于 RNAi 非同寻常的序列特异性和 dsRNA 给予途径的简单易行，使得 RNAi 成为研究那些已测基因组序列的生物基因组功能的首选。RNAi 已在秀丽线虫被证实是功能基因组研究的有力工具。

在几个系统中，RNAi 已被用作基因功能研究的强大工具，尤其在黑腹果蝇，RNAi 被作为快速评价抑制基因功能后果的标准方案。创建抑制秀丽线虫基因组的 19 000 个基因的 RNAi

文库计划正在进行。

在锥虫和植物中已建立转基因表达 RNA 发夹而产生可遗传的 RNAi。向非洲蟾蜍胚胎注射 dsRNA 可特异干扰靶基因表达，但 RNAi 在斑马鱼胚胎中仅中等程度有效，在这种情况下，特定的表型可被非特异效应干扰。显微注射 dsRNA 后，RNAi 在小鼠卵母细胞和预先移植的胚胎中有活性，并持续几个细胞分裂周期。

在研究那些仅有少许遗传工具可以应用的生物如奶虫螨、红粉甲虫等时，RNAi 的遗传分析尤其有用。RNAi 因而显著加快了对种属间形态和发育差异的比较进化生物学研究结果的解释。

用 RNAi 可以进行细胞分子的信号传导通路研究。

将来也许可以运用 RNAi 进行没有偏性的、表型驱动的遗传学筛选，例如可以通过构建反向的启动子系统产生随机化 siRNA 文库，从而利用 RNAi 进行反向遗传学筛选。

稳定表达的 RNAi 的发现具有广阔的应用前景。除了可以用 RNAi 稳定沉寂的胚胎干细胞株，并将其最终用于产生带某一沉寂位点的动物模型以外，更理想的是可以建立可诱导的 siRNA 系统（例如用 U6 启动子、四环素调节系统），用可调控的方式敲低那些对细胞存活必需的基因以研究其功能。同时，可以想像在不激活干扰素反应的情况下，将应用 RNAi 治疗人类遗传性疾病或病毒感染性疾病。由于 siRNA 的高度特异性，可仅用于致病突变转录本，例如 *P53* 或 *Ras* 等原癌基因转录本，而不影响野生型转录本。已经有用 RNAi 治疗 HIV 的尝试。

总之，序列特异的干扰基因表达的工具在功能基因组与生物医学中具有很广阔的应用前景，RNAi 无疑会成为这一研究领域先进的技术方法。

（赵　勇）

第十三章

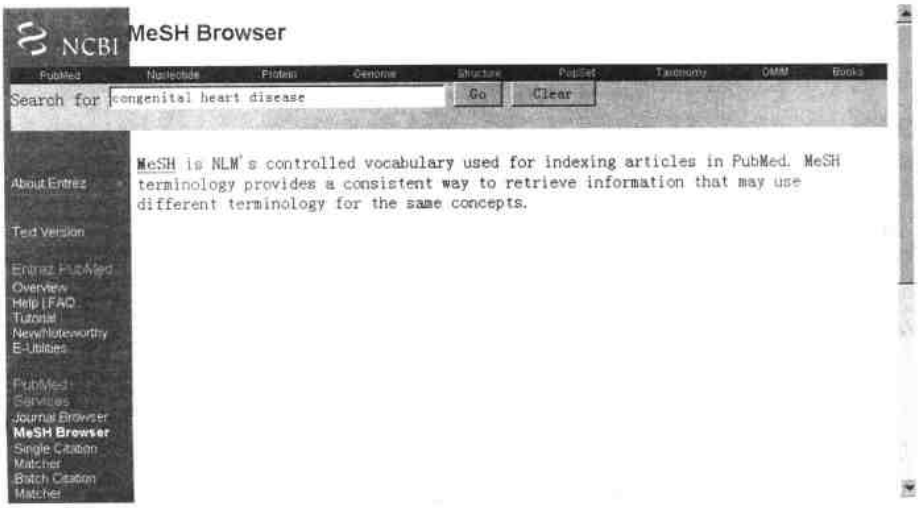
常用数据库及分子生物学软件

内 容 提 要

简要介绍了文献检索的方法、常用数据库及分子生物学常用软件。

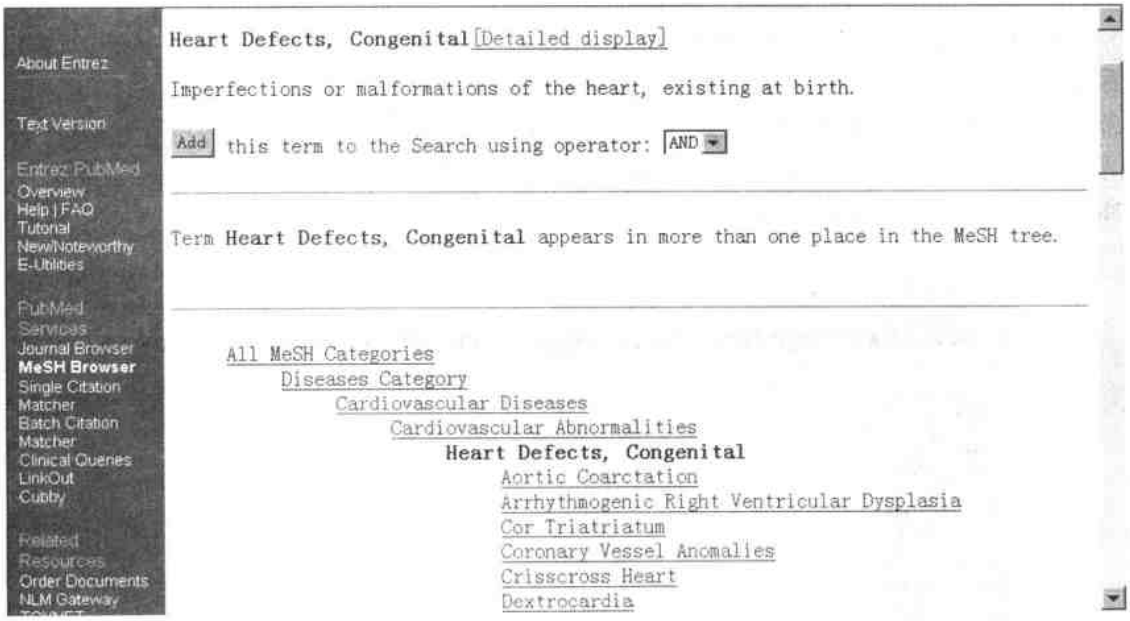
在科研工作中要随时通过阅读相关文献来了解相关领域的国内外研究进展。对生命科学工作者来说，收录有全世界约 4 600 种医学和生物学刊物的 MEDLINE(medical literature, analysis, and retrieval system online)数据库是文献检索的必去场所。目前 MEDLINE 含有的文献数已超过了 1 100 万。显然，想从这浩瀚的医学文献中准确地检索出所需要的文献必须借助于一定的工具。而由美国国家图书馆开发的 MeSH(medical subject headings)就是这样一个有力的工具。

MeSH 即医学主题词。由于医学科学的迅速发展，在现代生命科学文献中，同一个概念往往有不同的描述。这就给检索相关的文献带来了很大的不便。而通过 MeSH 术语则可以用规范的术语将相关的文献比较准确地寻找出来。在运用 MeSH 词之前，可充分利用 MeSH Browser 寻找欲检索关键词所对应的主题词。例如想了解“先天性心脏病”所对应的主题词，可进入 MeSH Browser，输入 congenital heart disease。



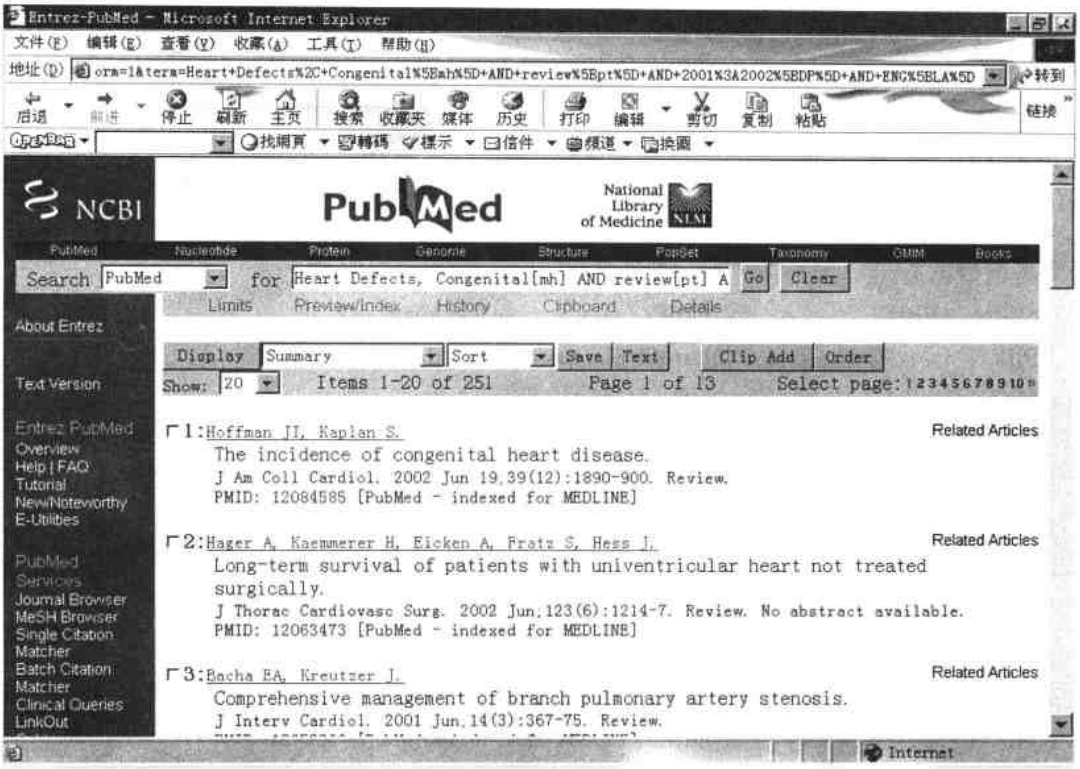
点击 go 后出现如下对话框：说明“congenital heart disease”不是一个 MeSH 词，它给了以下这些可能与之对应的 MeSH 词供选择。显然，这里应该是“Heart Defects, Congenital”。

选择该术语，点击浏览，则可以看到先天性心脏病的解释和该 MeSH 词的树形结构，先天性心脏病被归于心血管畸形中。



掌握了 MeSH 的使用，就可以进行比较复杂的文献检索。例如想了解一下近一两年有关先天性心脏病的最新进展，可以采用如下表达式：Heart Defects, Congenital[mh] AND review[pt] AND 2001:2002[DP] AND ENG[LA]，它的意思是查找从 2001 年至 2002 年用英文发表的有关先

心病的综述。PubMed 的查询结果共有 251 篇。当然还可以有更复杂的查询方式，例如查询某个著名研究机构发表的相关文献等等，这些都有待读者在实际工作中逐渐掌握。



上面简要介绍了文献检索，下面列出了一些常用的数据库及分子生物学软件，读者可根据不同的需要选择使用。

第一节　常用数据库

为便于初入门的读者进行研究的需要，现列出分子生物学研究比较常用的数据库。所有数据库的网址都经确认(2002/7/10)。

基因组数据库

GDB	http://gdbwww.gdb.org/
TIGR	http://www.tigr.org/
The Sanger Center	http://www.sanger.ac.uk/

核酸数据库

GenBank	http://www.ncbi.nlm.nih.gov/Genbank/index.html
EMBL	http://www.ebi.ac.uk/embl/

基因表达数据库

EGAD	http://www.tigr.org/tdb/egad/egad.shtml
Gene Expression Atlas	http://expression.gnf.org/cgi-bin/index.cgi
READ	http://read.gsc.riken.go.jp/
Gene Expression Omnibus	http://www.ncbi.nlm.nih.gov/SAGE/
TissueInfo	http://icb.mssm.edu/services/tissueinfo/query

启动子数据库

EPD	http://www.epd.isb-sib.ch/

蛋白数据库

SWISS-PROT	http://www.expasy.ch/sprot/
PIR	http://pir.georgetown.edu/
MIPS	http://mips.gsf.de/

蛋白基序和结构域数据库

SMART	http://smart.embl-heidelberg.de/
PROSITE	http://www.expasy.ch/prosite/
Pfam	http://www.sanger.ac.uk/Software/Pfam/
InterPro	http://www.ebi.ac.uk/interpro/
BLOCKS	http://www.blocks.fhcrc.org/
PRINTS	http://bioinf.man.ac.uk/dbbrowser/PRINTS/

蛋白相互作用数据库

BIND	http://www.bind.ca/
DIP	http://dip.doe-mbi.ucla.edu/
MINT	http://cbm.bio.uniroma2.it/mint/

疾病数据库

OMIM	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi? db=OMIM

代谢通路

KEGG	http://www.genome.ad.jp/kegg/kegg2.html
EMP	http://emp.mcs.anl.gov/

数据库综合查询及分析系统

SRS	http://srs.ebi.ac.uk/
Etrez	http://www.ncbi.nlm.nih.gov/Entrez/

细胞与组织培养数据库

ATCC	http://www.atcc.org/

文献查询及分析

PubMed	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi? db = PubMed
XplorMed	http://coot.embl - heidelberg.de/xplormed/

第二节　分子生物学常用软件

这里列出的分子生物学常用软件，基本上能满足平时科研工作的需要。

核酸低级结构分析

DNASIS 2.5

基于 MacOS 和 Windows 操作系统的基因综合分析软件，包括 DNA、RNA、蛋白质，优秀的限制酶分析软件，而且其在序列输出上非常完美，适于出版。

引物分析

Primer Premier 5.0

很好的引物设计软件，与之类似的软件有 Vector NTI Suite 7.1、DNAClub、Oligo 5.0 等。

综合分析软件

在目前所有综合分析软件中，GCG Wisconsin Package 可能是功能最为强大的一个序列分析软件，它提供了约 130 个程序。但其价格十分昂贵，适合于大的研究机构使用。而由 InforMax 开发的 Vector NTI Suite 7.1 是一适合 PC 机个人使用的、有名的综合性蛋白核酸分析工具包，具有常用的大部分分析功能，例如引物设计、多重序列、序列装配等，而且具有良好的数据库管理(增加、修改、查找)。同类的软件有 DNASTAR 5.1(又名 LaserGene)、Omiga 2.0 、DNASIS 2.5 等。

多重序列对排

许多综合性分析软件都有序列比对功能，如 Lasergene Suite、GeneDoc、Vector NTI Suite7.1 等，也有专门用于多重序列对排的软件如 MACAW 2.05、Clustal W 1.8(DOS 状态下操作)、Clustal X(Windows 界面，应用较广)等。其中 GeneDoc 的彩色输出非常漂亮，适合出版。

克隆策略绘图

Gene Construction Kit 2.0

优秀的质粒作图及克隆策略绘图软件，但操作比较难。其中较好的质粒作图的软件还有 Winplas 2.6 和 pDRAW32 1.0.33 等。

空间立体结构显示

RasMol 2.7.1 和 Swiss - PdbViewer 是功能强大的空间立体结构显示软件，但对非专业人员来说，使用比较困难。

序列格式转换

序列分析时，许多程序有各自的格式，如何将各种序列格式进行转换是件令人头疼的事。GeneStudio 出品的 Seqverter 1.55 是一很好的格式转换软件，它可以轻松地将序列转换为几种常用的格式，并将序列根据要求作一定修饰，但需注意的是该软件不具有序列排列功能等。

进化树分析

PHYLIP 3.6

由 Washington 大学 Joe Felsenstein 开发的免费软件 PHYLIP(穷人软件，作者 Joe Felsenstein 语)可能是最广泛使用的系统发育程序。它包含了大约 30 个程序的软件包，这些程序基本上囊括了系统发育的所有方面。对习惯 Windows 的读者来讲，似乎惟一不方便的是其需要在 DOS 界面下操作。

PAUP 4.0

PAUP 4.0 是另一个应用较为广泛的系统发育程序(对于已有 PHYLIP 3.6 这样优秀的软件，再去拿有限的科研经费去买这样一个商业软件，有点不值得)。

文献管理

Reference Manager 9.5

灵活的 Boolean 在线 PubMed 和其他各个数据库查询，也可查询本地资料库。文献管理方式十分灵活，可自动地将参考文献按不同期刊的要求排列，对经常著书立说的人来说，是一不可多得的好帮手。

(赵　勇)

第二部分

生物芯片技术

第十四章

生物芯片概介

内容提要

基因芯片技术是20世纪90年代发展起来的分子生物学技术，是各学科交叉综合发展的产物。它采用原位光刻合成或显微印刷技术在载体材料上形成DNA微阵列，与标记的样品核酸分子进行杂交反应。样品DNA/RNA在PCR扩增、体外转录的同时被标记，与微阵列杂交后经仪器扫描及计算机分析即可获得样品中大量基因的遗传信息。该技术可应用于新基因发现、高通量基因表达平行分析、大规模序列分析、基因多态性分析和基因组研究等，是基因功能研究的有力工具。常用的生物芯片分为三大类：即基因芯片、蛋白质芯片和芯片实验室。生物芯片的主要特点是高通量、微型化和自动化。芯片上密集排列的分子微阵列，使人们能够在短时间内分析大量的生物分子，快速准确地获取样品中的生物信息，极大地提高了传统检测手段的效率。本章还以名词解释的方式介绍了与芯片相关的分子生物学概念及实验方法，如杂交、突变、基因组的多态性、PCR、序列分析、SSCP等，以供备用和复习。

基因芯片技术是20世纪90年代发展起来的分子生物学技术，是各学科交叉综合发展的新产物。

在这个时期，最瞩目的科技成果是人类基因组计划的提前完成和后基因组计划的启动，它预示21世纪将是生命科学取得重大突破的世纪。在医学领域中疾病的基因诊断和基因治疗、干细胞克隆治疗、器官克隆、基因疫苗等医药高科技成果将深刻地改变人类的生活方式和健康理念，为此新技术和新方法的不断涌现将起到举足轻重的推动作用。生物芯片技术就在这种大背景下应运而生，它以高通量、集约化为特点，极大地简化和缩短了实验研究进程，它所体现的辨证统一的分析观为从分子水平向整体水平研究生命现象的回归架起了一座桥梁；它是可以和PCR技术媲美的分子生物学发展中的“里程碑”级的生物高科技，实用性极强，和疾病密切相关，尽管目前尚属起步阶段，已展现巨大前景。

第一节　生物芯片产生的科技背景及在医学中应用的优势

一、生物芯片的定义和特点

什么是生物芯片？

简单地说，生物芯片是指能对生物分子进行快速、并行处理和分析的邮票大小的固体器件。具体地说，生物芯片是指在一块大小不等(一般约 1～2 cm^2)的玻璃片、硅片、尼龙膜、凝胶或金属等载体材料上，以大规模阵列的形式排布了不同的生物分子探针(蛋白质、核酸等)，形成可与目的靶分子相互作用，并行反应的固相表面。将芯片与荧光标记的靶分子进行化学反应(如杂交、免疫反应等)，经过激发光扫描后，不同反应强度的标记荧光将呈现不同的荧光发射谱征，用激光共聚焦显微扫描仪或 CCD(charge couple device)相机收集信号后，经计算机分析数据结果，从而获得相关的生物信息。从另一个角度考虑，芯片分析实际也是传感器分析的组合，芯片点阵中的每一个反应单元都相当于一个传感器的探头，探测到的荧光信号被扫读之后，经图像分析和数据提取，再输出作进一步的生物学和统计学分析。

生物芯片的特征是什么？

生物芯片对样品的检测是以高通量、集成化、并行化和微型化为特征。举例来说，我们要研究肿瘤细胞的基因表达发生了哪些变化？或想了解在炎症时机体和细胞的反应性变化，就必须对肿瘤细胞的相关基因(如癌基因、抗癌基因、凋亡基因、细胞周期相关基因等)进行逐一分析；或对炎症相关的细胞因子、趋化因子、粘附分子等的表达状况进行研究。常规的做法是采用不同的实验技术在基因水平、mRNA 转录水平、或蛋白质翻译水平上进行研究。如果想观察 mRNA 表达水平的改变，首先要提取细胞的 RNA，然后通过逆转录反应，获得 cDNA，以其为模板进行定量 PCR 扩增，以确定不同基因的表达变化；也可采取 Northern 杂交的方法进行分析；不管怎样，总是一个基因、一个基因地、不连续地分析，效率之差显而易见。如果以一天分析一个基因的速度计算，100 个基因的分析也得花去 3 个月；加上不是并行性进行操作和分析，组间误差和批间误差难免产生。然而，用生物芯片能避免这方面的限制，芯片带给人们的最大优势是能同时分析和处理大量的样品。深入的生物学研究经常需要对数种样品的数据库作分析，例如一旦要观察多种生长因子在数种细胞系的不同时间点的表达状况的数据库进行分析时，利用芯片技术则可以特有的数据处理方式来展示现象之间联系。因为芯片在进行结果分析时，可将每种基因定义为 K 维空间中的一点(在这里 K 是被测样品数目)，基因之间在功能(表达)方面的相似性则可能揭示出它们在该空间中的依存关系，并以“簇”(Clusters)的方式表示，功能接近的可以聚类的方式归之为同簇；因而生物芯片特别适应于基因组和后基因组时代大规模生物信息分析的需要。

生物芯片的概念来自计算机芯片，因为其制作过程中采用了类似集成电路制作中的微加工技术，使得生物学实验中一系列不连续的过程如样品准备、生化反应及检测等步骤连续化和微型化，在一个小小的芯片上集成大量的生命信息。新近发展了“缩微实验室芯片(lab on a chip)”或简称为“芯片实验室”，它是在微芯片基础上建立的一种便携式分析系统，可集成式地完成样品制备、生化反应及结果检测 3 个步骤。也即通过微细加工工艺在芯片上制作微滤器、微反应器、微泵、微阀门、微电极等结构元件，以实现将不连续的化学分析过程移植到生物芯片中，像一个全功能的实验室，包含所有的分析平台；从样品制备、生化反应、功能检测到结果分析，一气呵成，从而极大地缩短了检测和分析时间，节省了实验材料，更突出体现生物芯片的连续化和自动化的优势和生物技术发展的方向。

二、生物芯片产生的科技背景

当我们今天在会议室、在休闲场所、在火车上，突然耳机铃声响起，就在此时、此地，就在你的手掌中，全部通话过程瞬间完成。这是 50 年前人们所不能想像的。正由于一场伟大的工业

革命——“硅革命”，才深刻改变了当今人类的生活。科学家们用硅作基质材料，将由无数的电子元件构成的庞大电路通过光刻加工，微缩到指甲盖大小的集成电路芯片中，使分离元件计算机的体积减少到笔记本大小。这场工业变革的发源地是美国的硅谷。

50年后的今天，生命科学领域的一场变革同样震撼人心。美国的硅谷又诞生了一类崭新的芯片——生物芯片，除了它所带来的巨大经济效益外，同样将深刻地改变着人类的生活。生物芯片模拟了集成电路芯片的工作原理，以大规模微阵列（microarray）的形式在面积为1～2 cm^2的硅片载体表面固相化排布了不同类型的、可寻址的生物分子探针。生物芯片和集成电路芯片一样，可以将个别的、不连续的化学反应单元及进行反应的支撑硬件都集合和微缩至硅片上进行。

生物芯片的影响将渗入到人类生活的方方面面。由于可从个人芯片化验室、环境监测芯片及药物反应芯片中高通量地获取详尽的资料，疾病的诊断和治疗变得个性化和即时化；远程诊断和远程医疗在家中都可进行；生物芯片技术还可以用于新药开发、农作物育种、微生物筛选、司法鉴定、食品卫生监督、环境保护等领域；甚至战争期间生物战迹也可通过芯片远距离收集。可以预测，随着生物芯片的使用、普及和扩展，人们将会感觉到，在我们的生活空间中，时间将变得更快、距离将变得更短、信息将变得更多、效率将变得更高。

自1991年Fodor等人提出DNA芯片的概念后，近年来基因芯片得到了迅猛发展。生物芯片技术的诞生是人类基因组计划（HGP，human genome project）启动、执行和完成的必然产物，也是生命科学进入后基因组时代的发展需要。随着人类基因组计划的基本完成，科学家获得和储备了人类及模式生物基因组DNA序列所包含的生物信息、具有高技术含量的生物材料和生物资源（如各种DNA或cDNA文库、遗传图谱及突变形式等），也为进一步深入研究人类遗传性疾病或肿瘤的诊断及治疗、开发和筛选新型药物的应用前景提供物质和技术的支撑。随着后基因组时代的到来，研究的重点也逐渐由发现新基因转向探索其功能，即面对大量的基因结构信息，如何对其功能进行大规模的研究，这是后基因组时代的主要研究目标。只有这样做才能真正体现HGP计划的价值。对不同个体、不同组织、不同发育阶段和不同生理状态下的基因进行表达差异的分析是完成蛋白质组计划的关键环节。面对人类基因组中几万个基因的复杂表达方式，对其相互作用和调控关系的研究是一项重要而艰巨的任务。而传统的建立在电泳基础上的基因表达、测序、突变和多肽性检测以及Northern－blot、Southern－blot等研究方法只能对有限的基因进行研究而且费时费力，很难在全基因组水平进行系统、宏观的研究。在这样的背景下，生物芯片技术应运而生，它能够以大规模、高通量、自动化的方式对基因进行研究。它不仅能利用基因组计划的大量信息在疾病诊断、药物筛选等领域发挥重要作用，而且还可为蛋白质组计划的实施提供大量新的线索。

三、生物芯片在医学中的应用概况与前景

（一）基因表达谱分析

利用cDNA文库和数据库可以制备cDNA微阵列芯片。一种基因在不同时间、不同组织结构、不同发育阶段及不同生物模式中的表达方式也不同。如果将不同个体、不同组织、不同细胞周期、不同发育阶段、不同病变、不同刺激（包括不同诱导、不同治疗阶段）条件下的细胞内mRNA或逆转录后产生cDNA与表达谱基因芯片进行杂交，可以对这些基因表达的个体特异性、组织特异性、发育阶段特异性、分化阶段特异性、病变特异性、刺激特异性进行综合的分析和判断，这为探明该基因的生物学功能和作用提供线索，并能迅速将某个基因或几个基因与疾

病联系起来，极大地加快这些基因功能的确立，同时进一步研究基因与基因之间的相互作用。

利用表达谱基因芯片研究基因功能的途径无疑是一种从分子水平向细胞和整体水平研究生命现象的回归，在整体中、在联系中观察个别基因的行为和功能是最客观和全面的分析方法。

目前，国内已有多家生物高科技公司制备了各类 cDNA 表达谱芯片，诸如癌基因和抗癌基因芯片、信号转导芯片、细胞周期检测芯片、细胞凋亡检测芯片、细胞因子、生长因子及生长因子受体检测芯片等，为基因表达的整体研究提供方便。

(二)基因突变检测

基因诊断是 20 世纪 70 年代末期伴随着基因重组技术而发展起来的一项以分析核酸中碱基序列特征为基础的新的诊断策略。

用生物芯片检测基因突变的一个成功的例子就是用 4 行砌砖式阵列(又叫叠瓦阵列，tiled array)检测人线粒体 DNA 的变异。检测时，所用的寡核苷酸(ODN)探针为 15 mers，也可以在 15 ~ 25 mers 的长度范围内。当这类突变检测芯片与带荧光的样品杂交时，可以在芯片中出现序列特异的杂交结果(详见第四章)。荧光测序结果很明显地区分出野生型及变异型的碱基序列。只要有 1 个碱基的突变，引起的错配可以使与点突变相邻的变异序列杂交信号明显减弱。

如将人线粒体 DNA 的全序列分为 16 000 个 20 mers 的寡核苷酸点在芯片上，制备成含 64 000个阵列的寡核苷酸突变检测芯片，只需作 1 次杂交就可检测出整个线粒体基因组中的变异。

目前这类突变分析芯片已成功地用于人酪氨酸酶基因、β - 珠蛋白基因、HIV - 1 蛋白酶和逆转录酶基因、囊性纤维变性基因、C - *ras*、C - *myc*、C - *erb* B2、*P53*、*Rb* - 1、*p*15、*p*16 等多种癌基因或抑癌基因突变的检测。理论上各种遗传病及肿瘤相关基因的变异，如血友病，杜氏肌营养不良症，地中海贫血，异常血红蛋白病，苯丙酮尿症的基因突变，都可以用相似策略的芯片鉴别出来。

(三)基因多态性分析

进行这类分析首先需要掌握有关待分析等位基因多态性序列方面的初步知识，然后选择适当的寡核苷酸探针和杂交条件。利用寡核苷酸芯片杂交技术有可能分辨只存在一个碱基差异的待分析 DNA 序列及差异序列的碱基顺序。

多态性分析最广泛的方法是用 ASOs 探针(等位基因特异的寡核苷酸探针)进行差示杂交。按下列步骤进行：①在玻璃片上制备 ASOs 平面点阵；②合成和纯化荧光标记的单链 PCR 探针；③标记的 PCR 产物与 ASOs 点阵杂交；④荧光扫描点阵杂交结果，鉴定基因的多态性。曾用这种方法测得人类酪氨酸酶基因第 4 个外显子的 5 个点突变位点。

基因组位点单核苷酸多态性分析(SNPs)由于其包含的信息量极大，加上平均每 1 000 个碱基就有一个单核苷酸多态性变异，因此近年来基于 SNPs 检测的基因芯片受到重视。曾报道了将凝胶测序法与高密度基因芯片技术结合，对 2.3 Mb 的人类基因组长度进行了 SNPs 筛选，最终确定有 3 241 个 SNPs 位点，其中 2 227 个位点被用来作为构建基因图的路标。

(四)序列测定

由 Sanger 创建的双脱氧核苷酸末端终止法和 Maxam - Gilbert 化学测序法的效率、可靠性差，不能满足大规模测序的要求。基于杂交原理的 DNA 芯片测序法已经建立，将芯片测序结果与凝胶测序方法互相配合或反复比较后，已经展示较好的前景。尽管目前用该方法还难以

对很大的未知 DNA 片段进行测定，但仍有发展前途。

测序芯片的工作原理是首先假设寡核苷酸的长度为 N 个碱基，在芯片上可建立 4^N 个长度为 N 碱基的随机序列的阵列集合，而待分析的靶序列的长度大约为点阵上寡核苷酸数目的平方根，例如一个含 8 个碱基长度的 65 536 个点阵的芯片，可测定约 200 个核苷酸的序列，可以此类推。

（五）药物基因组学（Pharmacogenomics）研究

药物学主要研究药物的制备、应用和效应；而药物基因组学则提供了基因组的全部序列及有关基因和功能的知识。从某种角度来看，这两门学科是互补和相通的，知道了基因组的全部序列，等于知道了被基因组编码的基因产物的全部序列以及提供了药物作用于有机体的有效靶标的序列，因为在蛋白质水平上，药物的作用就是要阻断或改变其靶蛋白功能。

由于芯片技术能在基因组的水平上能监控基因的表达，从芯片表达谱中所获得的信息很可能有利于研究药物的作用机制及药物对基因表达的调节作用，特别是一些小分子药物对基因表达的影响。同时芯片技术也在基因组和药物发掘之间架起桥梁，开拓了药物基因组学研究的新内容，包括先导化合物的筛选、药物的毒性、效应、最佳用量及药代动力学等。可以预言，芯片提供的信息与临床资料相结合必将加速和降低药物开发的成本和价格。表 14－1 概括了 DNA 微阵列在药物基因组研究中的特点、范围和优势。

表 14－1　药物基因组学中的微阵列分析内容

应　用	基因表达	序列分析
药物作用机制	表达谱的改变将提供有关药物功能的线索	再测序可以定位药物与蛋白质相互作用的部位
先导化合物鉴定	小分子的效应可以叠加在疾病表达谱的改变上	可使用基因分型的数据指导先导化合物的鉴定
先导化合物的优化	筛选对表达有最大影响的化合物	可将基因型数据与先导化合物的活性进行相关分析
毒理	表达谱的继发改变有助于毒理机制研究	毒性与基因型结合鉴定相关基因
转化	表达谱的改变有利于药代动力学研究	药物转化与基因型关联分析
药效	个体间表达谱的差异可决定药效	个体间多态性差异可以决定药效

（六）“中药基因组学”（TCM Genomics）

“中药基因组学”的含义是通过现代科学技术手段结合传统中药理论和现代科学理论，将中药的药性、功能及主治与其对特定疾病相关基因表达和调控的影响关联起来，在分子水平上用现代基因组学，特别是功能或疾病基因组学的理论来诠释传统中药理论及作用机理。

生物芯片在名贵中药材鉴定、药效分析、药物有效成分分析等方面也有极大应用潜力。香港理工大学在这方面作了开创性研究。见图 14－1。

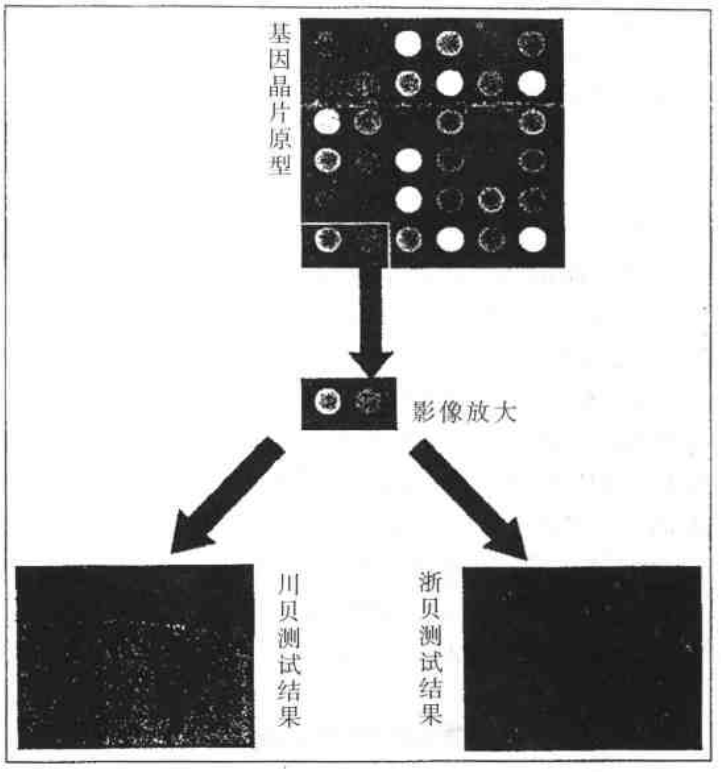


图 14－1　中药物种鉴定芯片

▲基因芯片鉴别川贝和浙贝示意图

(七)微生物菌种鉴定、致病机理及抗药性分析

DNA 芯片可根据细菌或病毒基因组分析结果进行基因分型、菌种鉴定及流行病学研究；还可以对微生物的重要基因，如毒力基因、抗药基因、致病因子等进行监测，寻找致病机制及耐药机理，为临床治疗提供依据。

分析药物作用下病毒基因表达情况，对了解药物作用机制及基因功能有很大的帮助。Kozal 曾利用高密度寡核苷酸芯片对 102 例病人中 167 株分离病毒进行研究，发现 HIV－1 的包膜 B 蛋白酶中氨基酸编码序列的多变性与该病毒抗蛋白酶抑制剂的抗药性有关。利用 DNA 芯片可以检测细微的致病微生物 mRNA 水平的改变，有助于从病原微生物—宿主基因组相互作用的角度探讨致病机理及治疗效果。

从基因芯片的研究发现 90%～95%的结核杆菌对利福平的抗药性与 RNA 聚合酶的 β 亚单位(*rpoB* 基因)的突变有关，包括错义突变、缺失突变等。利用芯片技术还可以对 2 个以上不同的抗药性基因(如 *kasA*、*pncA*、*emb* 等)同时进行杂交测定。

(八)克隆选择及文库筛选

将生物芯片与 DNA 文库和噬菌体文库结合，进行文库的分类和筛选，一般有下列两种类型：

1. cDNA 文库分类。将人体特定组织或器官的 cDNA 文库中的克隆扩增后，以微点阵方式

固定在杂交膜上，选用特定的探针系统与之杂交，再以计算机软件分析，就可快速、低廉、高效地对一个大容量 cDNA 文库中代表不同基因的 cDNA 克隆进行分类和分析，让人们更清楚简便地了解组织或器官在发生，发展，老化及癌变过程中基因表达水平，以及整个人体内基因网络的变化。

2.噬菌体文库分析。噬菌体抗体文库是噬菌体肽库中的一种，近年在免疫学和医学中得到广泛的应用，如分析未知抗原表位、抗原决定簇的定位分析、筛选抗体、寻找疾病相关的抗体、疫苗的研制及药物开发等。将目标抗原固定在生物芯片上，将所筛选的抗体文库加到缓冲液中，以液体流动形式不断地与固定化抗原作用。通过改变洗脱液的流速、收集不同时间点流出的噬菌体，通过动力学常数分析，就可最终筛选到高亲和力的小肽。

(九)监测蛋白质与蛋白质、蛋白质与小分子之间的相互作用

这类蛋白质芯片或肽芯片正在开发之中，但也面临许多问题有待解决，因为蛋白质之间的相互作用与杂交反应不同，后者只涉及线形序列之间的相互作用，而对于前者，由于蛋白质分子的三维折叠，分子之间的作用仅发生在肽表面，由此带来 3 方面的问题：①所制备的蛋白质芯片必须保持蛋白质的精细的折叠结构；而在芯片制备过程中使用的剧烈化学试剂、加热、干燥等反应条件很可能损害蛋白质的结构；②折叠蛋白质之间的相互作用比杂交反应表现出更大的序列依赖性，使得反应动力学和检测量更复杂；③荧光标记的蛋白质样品的获得还有待方法学的完善和提高。

第二节　生物芯片相关分子生物学概念

一、基　　因

基因(gene)代表一个遗传单位，在化学上指一段 DNA 顺序，该顺序可产生或影响某种表型，可由于突变而生成等位基因的突变体。

二、基因组和基因组的信息含量

基因组(genome)是指一个配子(精子或卵子)、一个单倍体细胞或一个病毒颗粒中所包含的全套基因。

基因组序列的组成和大小在不同有机体的形式和功能不同。一般而言，基因组的复杂度与有机体的复杂度成正比。细菌等简单生物基因组含有 $1\times10^6\sim5\times10^6$ bp；包括人在内的哺乳类生物约含 $1\times10^9\sim5\times10^9$ bp；像蠕虫或昆虫之类的基因组则含有 $1\times10^9\sim2\times10^8$ bp。基因组被分为不同的称之为染色体的节段，人含有 23 对染色体，其中 22 对为常染色体，1 对为性染色体；*E. Coli* 含有一个环状染色体；酵母含有 16 个线性染色体。表 14－2 列举的是基因组中所包含的信息含量。

来自于同一有机体的所有细胞含有相同的基因组，而同种不同个体的基因组约有 99.9% 的同源性及 0.1% 的多态率。

表 14－2　基因组包含的信息含量

信息	DNA	RNA	蛋白质
遗传的蓝图	全部	部分	部分
基因组表现度	100%	3%	3%
类型	编码和非编码	编码	编码
形式	双链	单链	折叠
组成	核酸	核酸	氨基酸
杂交	可以	可以	不能
碱基	A,T,G,C	A,U,G,C	没有
基本组成单元	4	4	21
转录效应	无	有	有
翻译效应	无	无	有
生化修饰	甲基化	多聚腺苷酸化	磷酸化
微阵列测定	可以	可以	暂无

基因组分析能提供的信息包括：基因拷贝数、染色体的数目及 DNA 一级结构中单个碱基的差异。但是在 DNA 水平上的基因组分析并不提供任何有关基因表达量的信息；而在 mRNA 水平上的基因组分析却可为基因表达提供测量手段；基因组的信息指令最终是以蛋白质的活性体现出来的。

三、真核基因组和原核基因组

真核基因组和原核基因组结构的比较特点主要有以下几点：

1. 真核细胞具有典型的细胞核结构，整个基因组分布在各个染色体中，染色体组装在核内，外面有核膜包裹，这就决定了真核细胞的基因表达必然受到其高级结构的影响转录和翻译两个过程，不能同时在同一空间进行，从而可以在更多的水平上进行调节控制，这反映了由低级向高级的一种进化。

2. 原核基因组具有操纵子结构，一群功能相关的结构基因（多顺反子）串联在一起，连同它们前面的调节区 P－O 区（P 为启动子，O 为操纵基因）共同组成操纵子（operon）；数个操纵子还可以受一个共同的调节基因 R（在乳糖操纵子中用 I 表示）控制。调节基因可与操纵子相连，也可以远隔一段距离。但这种操纵子结构模式尚未在真核基因组中发现，真核基因转录物是单顺反子，即一个结构基因经过转录和翻译分别生成一条 mRNA 和一条多肽链。

3. 原核和真核细胞中的结构基因中均无重叠现象，即同一部分 DNA 顺序不编码两种蛋白质，但是真核基因是不连续的，具有内含子和外显子结构，称之为断裂基因；而原核基因是连续的，不含内含子结构。

4. 真核基因组内不编码的序列大于编码的序列，人类基因组内编码顺序 < 5%，而在低等的病毒基因组中，95% 以上为编码序列，在细菌的基因组中，编码顺序可达 50% 左右。

5. 真核生物 DNA 中含有大量重复顺序，重复频率从数次到几百万次，各种重复序列的长度及重复次数不同，其功能也有极大差异。原核生物 DNA 顺序基本上没有或极少有重复顺序。

四、基因表达

基因表达是指以基因 DNA 为模板，合成信使 RNA（mRNA）和蛋白质的过程。基因表达的第一步是基因的转录，每一个基因 DNA 中能转录加工为 mRNA 的部分或能提呈为 mRNA 的部分，称之为该基因的编码序列。人类基因组中大约 3% 的部分可以编码序列的形式存在。基因表达的第二步则为 mRNA 的翻译，它将产生有生物活性的蛋白质，于是基因组中所隐含的功能指令全都得到了体现。

基因表达是遗传因素和环境因素相互作用的结果。遗产因素指启动子、增强子、剪接位点等转录调控区的 DNA 序列，决定基因表达活性；而环境因素包括温度、应激、光及其他的信号等将导致激素或其他信号分子的改变，从而影响基因的表达水平。因此，对 RNA 水平进行分析不仅提供机体遗传性方面的信息，还可了解机体功能状态的动态改变，例如在热休克、药物治疗、代谢或疾病状态时，mRNA 的水平均发生改变。有人提出假设，认为多数或所有的人类疾病都伴有基因表达的改变，在整个基因组的水平上监控基因表达的改变已引起商界的极大兴趣，也为 cDNA 表达谱芯片的研究和发展提供基础。

五、基因突变

从 DNA 到蛋白质的过程十分复杂，要确保每一步的精确进行，如果 DNA 结构发生改变就称为基因突变（gene mutation）。基因突变是产生新基因的源泉。从群体上讲是生物进化的动力，但对个体来说，大多数情况下都是有害的，或造成突变个体的死亡；或引起生物的疾病。

造成基因突变的原因很多，大致有以下几点：

1. 复制误差。细胞复制中难免出现错误，如胞内的 DNA 修复系统不能及时发现并修复，该误差将会在下一次的复制中遗传下去形成突变。

2. 取代突变。指 DNA 分子中的一种碱基被另一种碱基替换。如果是同一类碱基替换（如嘌呤取代嘌呤，嘧啶取代嘧啶）则称为转换（transition）；若属嘌呤与嘧啶之间的互换，则称为颠换（transversion）。这类点突变对蛋白质功能的影响是不同的。可为中性突变（neutral mutation），一般碱基的改变发生在三联密码的最后一个，由于密码子的简并性而不改变原有的氨基酸种类；也可为错义突变（missense mutation），碱基的改变引起氨基酸的改变。如果新的氨基酸与原来的性质相近或处于蛋白质的非功能区，则对蛋白质功能影响不大；否则将影响蛋白质功能或改变蛋白质构象，使功能减弱或丧失。第三类取代后果则为终止突变或无义突变（nonsense mutation），碱基突变形成终止密码子，使其下游的遗传密码子不再翻译，蛋白质合成终止。

3. 缺失（deletion）突变和插入（insertion）突变。无论是哪一种，都将改变 DNA 序列的大小和位置，如果同时改变了原来 DNA 编码序列的阅读框架（reading frame），那么合成的蛋白质完全错误，例如杜兴－别克肌肉萎缩症就是由于缺失突变造成的。

4. 染色体易位（chromosome translocation）或基因重组（gene recombination）。这类突变不限于单个基因，大范围的突变会造成染色体的形态异常和新的基因组合。

六、基因多态性

DNA 分子中，单个碱基的多态性是普遍存在的，它是人类个体之间千差万别的物质基础。如果 DNA 分子中单个碱基的改变是遗传下来的，通常称之为多态性（polymorphism）；而在有机体生命过程中所获得的单个碱基的改变则称为突变。若以人基因组 DNA 含 3×10^9 bp 及

0.1%的单碱基多态性频率来计算,那么任何两人之间,可找到300万多态性碱基差异。单个碱基的改变如发生在蛋白质的编码区,多态性的出现可能会改变蛋白质序列甚至基因产物的生物化学活性。基因产物的功能涉及人类健康的方方面面,因此多态性的检测必将成为功能基因组学的中心问题。

DNA水平上的多态性可能导致同种异体基因表达的差异。由于mRNA是DNA编码区的真实拷贝,故mRNA分析可鉴定DNA编码区的多态性。

根据DNA多态性的差别和影响,可将其分为:

1.DNA位点多态性(site polymorphism)。位点多态性是由于等位基因之间在特定位点上DNA序列存在差异,这些位点上某一碱基的存在与否将决定这段DNA能否被某一限制性内切酶水解,从而获得大小不等的片段,以分析个体间DNA的位点差异。

2.长度多态性(length polymorphism)。由于相同的重复序列重复次数不同所致,也称为可变数目串联重复序列(Variable Number of Tandem Repeats, VNTRs),这种多态性在人群中频率极高。非编码区内的串联重复序列通常存在于间隔DNA(spacer DNA)和内含子内。其中的串联重复单位从最短的2 bp起,长短不等,重复次数可少至数次,也可多至数百次。这类重复序列是组成卫星DNA的基础。卫星DNA可分3类:大卫星DNA(Macrosatellite DNA)、小卫星DNA(Minisatellite DNA)、微卫星DNA(Microsatellite DNA)。微卫星DNA是一类更简单的寡核苷酸串联重复序列,重复单位为2~4 bp,重复次数10~60次左右。常见的是以$(AC)n$和$(TG)n$的二聚核苷酸为重复单位。除了在内含子、间隔DNA中存在外,微卫星DNA也存在于编码区,按照三联密码的特点,这些编码区的重复序列必是3 bp组成的重复单位,如$(AGC)n$存在于人雄激素受体基因内。

一般而言,功能基因片段由于受选择压力的影响,表现出高度的保守性。杂合度低,突变率也低,表现在蛋白水平上平均变异率为20万年中每400个氨基酸才发生一个氨基酸变异。相反,基因组中DNA重复序列则不受选择压力的限制。这表现为DNA重复序列广泛分布于整个基因组内,并以多样化形式存在。任何DNA重复序列的变异都大于单拷贝的编码DNA序列,其突变率也高于编码顺序。所以,DNA串联重复序列在个体间最具有变异性。

七、DNA的变性与复性

当DNA双螺旋互补碱基对之间的氢键被破坏,分离成两条单链的过程,称为DNA变性。加热和加碱均能使DNA变性(denaturation)。升高温度使双链解开的过程又叫熔解,能使两条链分开一半时的温度称为熔解温度,即常说的Tm值。不同的DNA分子由于分子中G/C或A/T碱基的比例不同,它们的解链温度也不同。解链以后的单链DNA,在不受任何干扰的条件下,若缓慢地降低温度,分开的双链上,互补碱基间还可以再配对形成氢键,从而使两条单链DNA分子恢复成双链DNA。这种逆向的过程就叫DNA的复性(renaturation)或退火(annealing)。DNA分子间或DNA分子与引物之间的变性和复性过程是基因分析技术的重要组成步骤。

八、探　　针

基因探针(probe)是用于探测内源或外源靶基因的一段特异性的核酸序列,它可以是DNA或RNA;也可以是某一基因的全长或其中的一部分长度。一般可将探针分为基因组探针和cDNA探针两大类。前者是指由基因组序列中分离出来的DNA片段;后者则为经RNA逆转录出来的互补DNA片段。两者不同之处显然在有无内含子序列。

作为探针，必须具备的基本条件为：①能特异性（或选择性）地与待检测的靶分子结合；②探针分子本身必须带有一个显示标记（marker），如放射性同位素、荧光素、酶标记抗体等，以便能显示和证明探针与靶分子的特异性结合。探针的工作原理是基于核酸分子之间的配对杂交。检测前首先将探针分子和靶分子变性成单链，如果探针的一条链将与被检测 DNA 的一条链完全互补结合，说明被测基因与探针一致或检测物中存在与探针序列相关的基因序列。

九、杂　　交

杂交（hybridization）是指两条互补的单链核酸形成一条稳定的双螺旋分子的过程。杂交反应可以发生于溶液中两个互补分子之间；而一个溶液中的分子与另一个被固定在支持物上的分子之间也能发生杂交反应。

杂交过程大致如下：检测前首先将探针分子和靶分子加热变性成单链，将待检测的单链靶核酸固定在适当的载体上，同时探针分子也必须做好示踪标记，然后将两者置于杂交反应体系中进行反应（固－液杂交）。如果探针的一条链将与被检测 DNA 的一条链完全互补结合，说明被测基因与探针一致或检测物中存在与探针序列相关的基因序列。

杂交方法见表 14－2。

表 14－2　主要核酸杂交方法比较

名　称	靶分子	探　针	优　点	缺　点
Southern	DNA	DNA/cDNA	确定分子量	电泳＋转移
印迹杂交				操作繁琐
Nouthern	RNA	RNA/cDNA	确定分子量	电泳＋转移
印迹杂交				操作繁琐
Dot blot	RNA	RNA/cDNA	简单、快速	不能确定分子量
（斑点杂交）			大量	
Slit blot	RNA	RNA/cDNA	简单、快速	不能确定分子量
（狭缝杂交）			大量	
In situ	RNA	ODN/RNA/cDNA	不需提取核酸	组织细胞形态完整
Hybridization				
（原位杂交）				
RNA 酶保护实验	RNA	RNA/cDNA	mRNA 定量	mRNA 末端定位
（RNase protection assay）				内含子定位
核酸酶 S1 保护实验	DNA/RNA	RNA/cDNA	转录起始点分析	内含子剪切点分析
（NucleaseS1 protection assay）				

DNA 芯片测定就是利用荧光标记的单链分子与芯片表面的单链序列之间的杂交反应而进行的。芯片微阵列工业就是建立在杂交这个生物化学过程的基础之上的。

十、PCR

PCR(Polymerase Chain Reaction)又称聚合酶链反应,是一种在体外(试管、切片等)扩增核酸的技术。该技术模拟体内天然 DNA 的复制过程,利用 DNA 聚合酶依赖于 DNA 模板的特性,在附加的一对引物之间诱发聚合酶的合成反应。该技术由高温变性,低温退火及适温延伸 3 步反应组成一个周期,反复循环进行,最终使目的基因扩增的产量(即拷贝数)以指数方式上升。在 n 个循环后,被扩增的 DNA 产量为 2^n 拷贝。如 30 个循环后($n = 30$),反应终产物的拷贝数为 2^{30},也即增加 10^9 个拷贝,实际浓度增加了 $10^5 \sim 10^7$ 倍。

组成 PCR 技术每一循环周期的 3 个主要步骤是:

1.变性(denaturation)。待扩增的双链 DNA 分子,于高温下(94 ~ 95 ℃)解链成为两条单链模板的过程。

2.退火(annealling)。指人工合成的两条特异的寡核苷酸引物在低温(50 ~ 65 ℃)条件下,分别与各自的目标 DNA 片段两则的两条单链互补结合的过程。

3.延伸(extension)。DNA 聚合酶在适温下(70 ~ 75 ℃)将 4 种 dNTP 从引物的 3′端(3′ - OH)开始掺入,并沿着 DNA 模板由 5′→ 3′方向延伸,合成新生的 DNA 单链。

从上所述,PCR 扩增效率的指标为特异性、高效性和忠实性,该技术的主要特点为:

1.操作简便和省时。

2.灵敏度高。模板的加入量约为 10^2 拷贝的靶序列,就可以进行扩增反应。从理论上讲,一根头发,一个精细胞,一个二倍体细胞内的 DNA 都可得到扩增。

3.由于 PCR 极大的扩增效率及灵敏度,样品极微量的污染便可导致假阳性的出现。

4.特异性强。由于 DNA 聚合酶反应只是在一对引物之间引发,这就决定了扩增片段的特异性。

5.模板材料易获得。细菌和病毒标本、血液或细胞标本化石、鼻、咽、口腔及生殖器的拭子标本等,甚至经固定和包埋的石蜡或冰冻组织标本或 DNA 已有损伤或降解的古化石标本均能进行 PCR 扩增反应。

6.碱基错掺。由于 *Taq* DNA 聚合酶酶无 3′→5′外切活性,对错误掺入的单核苷酸无校正修复功能,故 PCR 反应产物会有一程度的碱基错掺,一般可以忽略不计的。但在进行 PCR - Cloning 实验时,必须考虑错掺是否会引起阅读框架的漂移或改变。

7.PCR 扩增终产物 3′端 - OH 的加尾。随机加上的单核苷酸,形成“毛边”碱基,以 A 的加入具有优势。在进行 PCR 产物克隆时必须考虑这种可能性。分析 PCR 产物两单链 3′端的碱基,发现总带有一非模板依赖性的碱基,而这种多出来的几乎总是 A。利用这一特性,人们构建了一类新型的分子克隆载体——即 3′端 dT 或 ddT 加尾载体,统称 T 载体,它可与末端携带 dA 的 PCR 产物进行连接,简化克隆步骤。

8.引物的 3′端。延伸将从 3′端 - OH 引发,故不能进行任何修饰。

9.引物的 5′端。由于 5′端的作用仅为限定 PCR 产物的长度,对扩增特异性影响不大,因而可被修饰。例如加限制性酶切位点 引入列启动子序列(如 T7 启动子)、插入与缺失突变序列、标记生物素 / 荧光素(FITC 或 FAM 等)/ 地高辛 / EDFA—Fe^{3+} 等、引入蛋白质结合 DNA 序列(顺式元件)。

第三节　生物芯片的分类简介

和所有的分类方法一样，同一种芯片根据不同的标准，就有不同的归类组别和定位。最初的生物芯片，用于 DNA 序列的测定、cDNA 表达谱分析、基因突变的检测和鉴定等，故称之 DNA 芯片或基因芯片。目前这一技术已扩展至免疫反应、受体结合、小分子捕获、药物筛选等领域，种类繁多，分类也趋于详细和全面。下面将作一简单介绍。

1. 根据基片上交联固定的识别分子种类不同，可分为基因芯片、蛋白质芯片、肽芯片、细胞芯片、组织芯片、寡核苷酸芯片等。

2. 生物芯片的结构特征、分析过程来分，可分为以亲合结合技术为核心的微阵列芯片和以微管道网络为结构特征的微流控芯片。

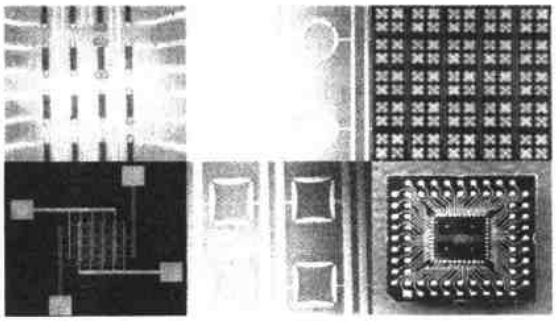


图 14－2　用于基因表达、蛋白分析及药物筛选的各类主动式芯片

3. 根据芯片的功能又可分为测序芯片、基因作图芯片、基因表达谱芯片、克隆选择和文库筛选芯片、物种鉴定和基因组分型芯片、突变检测芯片、多态性分析芯片、药物芯片、噬菌体文库筛选芯片等。特别值得一提的是能将样品分离、纯化、提取、混合、反应、检测等功能集合于一体的微缩实验室芯片(lab on a chip)，将代表微流路生物分析技术发展的最高水平。

4. 根据芯片的用途，还可分为分析芯片、检测芯片和诊断芯片。

5. 根据生物化学反应过程分类，可将生物芯片分为不同的类型：例如用于样品制备的生物芯片，生化反应生物芯片及各种检测用生物芯片等。

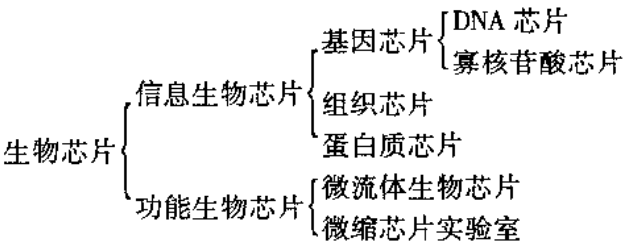
(1)样品制备芯片。其目的是将通常需要在实验室进行的多个操作步骤集成于微芯片上。目前，样品制备芯片主要通过升温、变压脉冲以及化学裂解等方式对细胞进行破碎，通过微滤器、介电电泳等手段实现生物大分子的分离。

(2)生化反应芯片。即在芯片上完成生物化学反应。例如，在芯片上进行 PCR 反应的 PCR 芯片(图 14－3)。由于检测和分析的灵敏度所限，通常在对微量核酸样品进行检测时必需事先对其进行一定程度的扩增，所以用于 PCR 的芯片无疑为快速大量扩增样品多个 DNA 片段提供了有力的工具，还可以节约实验试剂，提高反应速度。

(3)检测芯片。用来检测生物样品的。例如用毛细管电泳芯片进行 DNA 突变的检测，用于表达谱检测、突变分析、多态性测定的 DNA 微点阵芯片(也称 DNA 芯片)；用于大量不同蛋白检测和表位分析的蛋白或多肽微点阵芯片(也称蛋白或多肽芯片)。

国内有的单位曾将所有的生物芯片按照以下方法进行分类，也能很好地反映生物芯片的

全貌，特介绍如下，以供参考（其中并没有将药物芯片单独归类）：



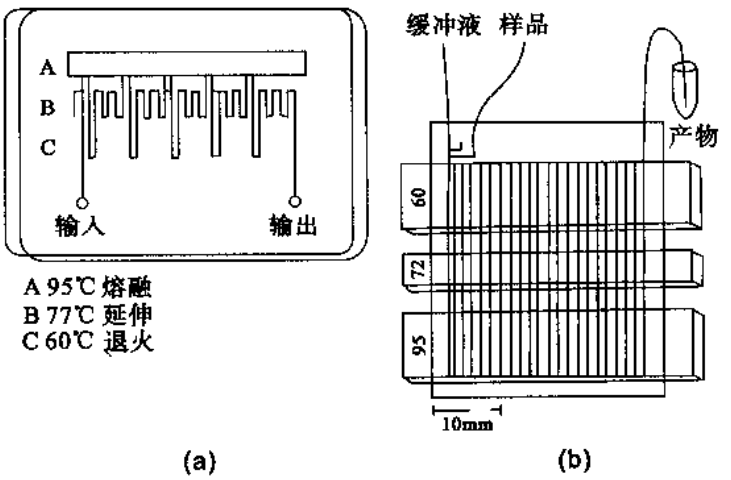


图 14－3　PCR 芯片示意图

(a)PCR 芯片；
玻片(50 mm × 40 mm × 0.55 mm)
蚀刻 90 μm × 40 μm 槽硅烷化盖片(0.55 μm)
有进出液孔(400 μm)环氧封装石英管
(内 φ100 μm)
微量注射泵输入 PCR 混合物及缓冲液

(b)加热用恒温模块
扩增 20 循环，50 min
扩增液电泳分析

（杨渝珍　曹慧青　李　健）

第十五章

生物芯片的制备和检测

内容提要

基因芯片技术是众多学科和技术相互渗透融合的产物。其产生背景是人类基因组计划的实施和发展。将成千上万种核酸探针分子有序的排列在面积不大的载体材料上并与标记的样品核酸杂交，通过放射自显影或荧光扫描显示，可一次获得大量有用的生物信息。因而可以进行宏观、系统、高效的研究。基因芯片的相关操作包括：基因芯片的制备、靶基因的制备、杂交和检测，其优势在于快捷、高效。但需建立芯片制备和使用过程中的标准化程序，提高其准确性和可靠性。基因芯片技术是对基因组功能进行高通量研究的有力工具。已经证明可用于基因发现、基因表达研究、杂交测序以及多态性检测等各领域。随着生命技术、电子技术以及计算机软件等的不断向前发展，基因芯片技术本身会逐渐被完善，也必将得到广泛的应用。

第一节　载体的表面修饰

一、芯片载体和载体材料

芯片的核心技术在于在一个有限的固相表面上刻印大量的生物分子（DNA 、蛋白质）点阵，故把用于连接、吸附或包埋各种生物分子，并使其以固相化的状态进行反应的固相材料统称为载体或基片。一种理想化的载体除了能有效地固定探针外，还必须允许探针在其表面与目标分子能稳定地进行杂交反应。

生物芯片所用的载体大多数是固体片状或者薄膜类，制作生物芯片的材料必须具有以下的基本特征：①具有良好的光学性质，能适应透射光和反射光的测量；②载体表面必须具有可以进行化学反应的活性基团，以便能与生物分子或修饰分子进行偶联；③使单位载体上结合的分子数达到最大的容量；④载体应当是惰性的并且具有足够的稳定性，包括机械的、物理的和化学的稳定性，而惰性是指载体所具有的特异性吸附或其他性能都不干扰生物分子的功能，而稳定性是在进行芯片制备和分子杂交过程中，能抗一定的压力和酸碱变化；⑤载体具有良好的生物兼容性以利于制作不同种类的生物芯片。研究者应根据具体的实验情况来选用适合于自己的片基材料以达到最好的效果。

制作生物芯片的载体材料很多，其理化性质符合要求的超过百种，它们大致分为4类，即无机材料、天然有机聚合物、人工合成的高分子聚合物、高分子聚合薄膜。具体而言，有普通玻片、硅片等钢性支持物和聚丙烯膜、硝酸纤维素膜、尼龙膜等薄膜型支持物。另外，还可用聚丙烯酰胺作为支持介质，将聚丙烯酰胺凝胶固定在玻片上，然后将合成好的不同探针分别加到不同的胶块上，制成以凝胶块为点阵的芯片。由于蚀刻技术的发展等原因，塑料已被视为将来制作芯片的理想材料。聚二甲基硅烷(PDMS)是现在最常用的塑料基底材料，它具有以下优点：容易塑造，光学透明，化学惰性与稳定性等。但是目前常用还是玻璃片、金属片，高分子材料薄膜等，其中以玻璃片应用最为广泛。

1.膜。与玻片、硅片相比，膜的优点是与核酸亲和力强，杂交技术成熟，通常无需另外包被。由于尼龙膜与核酸的结合能力、韧性、强度都比较理想，以膜为基质的芯片绝大多数采用尼龙膜。Schleicher & Schuell 公司是专业生产杂交/过滤膜的公司，著名的硝酸纤维素膜即是该公司首家推出的。S&S 的 Nytran SuPerCharge 生产的正电荷尼龙膜带电量是常规正电荷尼龙膜的3倍，与核酸的结合力更强，而且又大大降低了背景和非特异结合，克服背景高的缺点，加上韧性强，反复杂交10次依然保持表面平整的优点，成为制作尼龙膜芯片的最佳选择。

2.玻璃片。玻璃片是应用最广泛的，主要原因是：①材料来源方便且价格便宜，表面羟基可以用 N,N－二乙氧基氨丙基三乙氧基硅烷作表面处理，然后能够偶联核酸、酶分子、抗体(抗原)或其他多肽蛋白质分子，并防止其在杂交洗涤过程中被冲洗掉。经表面化学处理的玻片是一种持久的载体，它可耐受高温和高离子强度。也可以用巯基标记的生物分子直接和玻璃表面作用等。但是，往往价格低廉的玻片表面不均匀，这种异质性会导致表面活化的问题从而影响芯片的质量，现在高质量的底物玻片可以从下面的公司获得(http://arrayit.com 和http://www.cmt.corning.com)，这种芯片在表面活化中具有很好的稳定性和均一性。②大多数生物芯片采用了发光检测的方法，不管投射光或反射光它均适合。玻璃芯片可使用双荧光甚至多荧光杂交系统，可在一个反应中同时对两个以上的样本进行平行处理。③可以用光刻的方法在玻璃芯片上刻出微电路，或者它与刻好微电路的橡胶压紧成联合使用的微电路，通过微量注射泵在载体上进行生物分子的联结，DNA 杂交，清洗和检测等。④其他：如玻片具有不浸润性，使杂交体积降低到最小，因此提高了退火时的动力学参数；而且疏水表面可以使点密度大于亲水表面(因为亲水表面上样品点将扩散)；玻片的荧光信号本底低，不会造成很强的背景干扰。相对玻片而言，另一种载体硅片除了以上优点外，其均质性更好。

S&S 公司的 CAST Slides(Cat.No.10484181,20/box)将 SuPerCharge 正电荷尼龙膜附着在玻片上，这种特殊的玻片综合了尼龙膜的高亲和力和玻片刚性的优点，是世界上第一个膜结合玻片。而 FAST Slides(Cat.No.10484182,20/box)则是在玻片表面包被一种专利的聚合物，这种聚合物能迅速与 DNA 以非共价但是不可逆的方式结合。由于表面包被层的多孔性和厚度使单位面积的 DNA 结合能力比常规化学表面处理玻片要高得多，使得检测更加灵敏。其杂交方式和传统的杂交一样。适用的检测方法包括同位素检测、化学发光法和荧光检测；由于包被的多聚物有效降低对入射光的散射，FAST Slide 同样适合用激光共聚焦成像系统进行荧光检测。

3.硅片。硅是非常常用的一种材料，这是由于在微电子领域，人们已经摸索出了一整套成熟的硅加工工艺。对于需要温度控制的生物反应，例如聚合酶链式反应(polymerase chain reaction，PCR)，链置换扩增(strand displacement amplification，SDA)，由于硅具有良好的导热性，更是成为人们的首选。但是硅的缺点是不透明，不利于光学检测，并且具有一定的导电性，尤其是具有比较强的表面非特异性吸附。因此在制作毛细管电泳芯片时，人们一般不选用硅，而选用

玻璃或者塑料。由于在使用生物芯片时，必须考虑到生物相容性问题，而玻璃、塑料表面所具有的各种功能基团使其很容易进行化学修饰，从而使其生物兼容性大大提高。因此有许多种芯片是用玻璃或者塑料制作的。

基因芯片的英文名称表示法很多，有 Gene Array、GeneChip、DNA chip、microarray 和 macroarray 等等。在了解芯片的载体材料后，就不难理解为什么习惯于将 DNA chip、GeneChip、microarray 多指点阵密集度很高的玻片基质的微阵列（有时也用来指点阵密集度很高的胶膜芯片）；而 macroarray 则多指点阵密集度较低的尼龙膜基因微阵列。其中，GeneChip 是被 Affymetrix 公司注册的专利名称，特指该公司的产品系列。

二、载体化学修饰

载体化学修饰即载体的活化。为使探针稳定地固定于介质表面，需对介质表面进行化学预处理。活化试剂 EDC、NHS 等通过化学反应在载体表面上键合上活性基团如氨基、环氧化物、巯基、醛基、肼基，以便于与配基相结合，形成具有生物特异性的亲和表面，用于固定蛋白质、核酸、多肽等生物活性分子。通常在钢性载体表面必须衍生出活性基团如羟基或氨基（视所要固定的分子为核酸或寡肽而定），薄膜型载体为使其表面带上正电荷以吸附带负电荷的探针分子，通常需包被以氨基硅烷或多聚赖氨酸等。

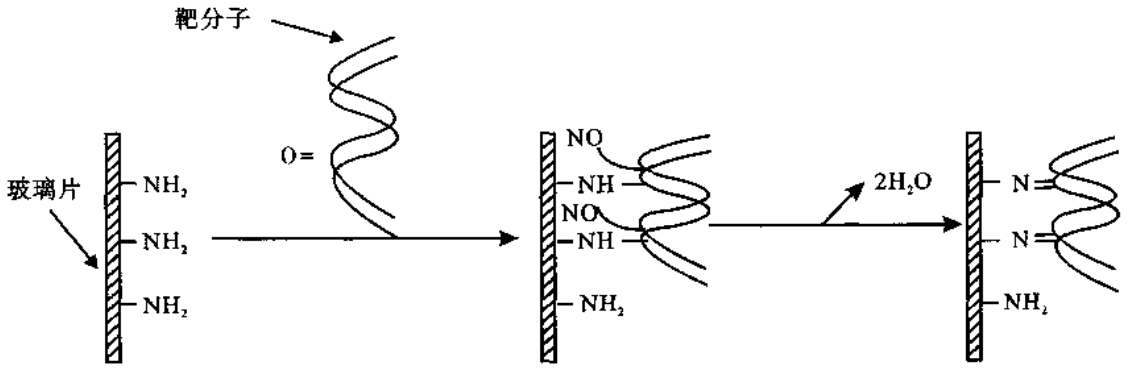
由于片基的表面修饰主要以化学修饰为主，面化学修饰的基本原则是：在片基表面上必须存在合适的功能性化学基团以便偶联生物分子，因为未经修饰的玻片是不会键合生物分子的，这种偶联将阻止刻印点上的靶分子在杂交和漂洗过程中脱落或扩散；并且提高刻印表面的质量。芯片载体处理的基本要求为：①修饰后的表面是均匀而且硬质的；②易于对修饰后的表面进行共价的或非共价的化学修饰；③在荧光探针激发波长范围内有一个低的背景荧光；④具有较高的信噪比（signal - to - noise ratio，S/N）；⑤允许有效的固定各种不同的生物材料；⑥允许在微表面上均匀地进行大量和高密度的样品点阵或刻印；⑦在同一个芯片上所有点的变量相关系数（Coefficient of Variation，CV）是很低的，一般低于 10%，这样能够保证在芯片的任何位置上均能进行有效的刻印。

目前，有很多方法已被用来对载体基片进行表面化学处理，制备成多聚赖氨酸修饰载体、醛基修饰载体、氨基修饰载体、巯基修饰载体、琼脂糖包被的载体、丙烯酰胺硅烷化载体、telechem 硅烷化玻片载体等不同类型。其基本过程都是把玻片的硅表面根据不同的要求而活化成不同化学活性的表面，这种过程往往在普通实验室条件下即可进行。选择的原则是依据片基的表面特征和将固定的生物配基分子的化学特征，例如表面为羧基的载体，用 EDC 和 NHS 把表面变成活泼的酯基，然后再与不同的生物分子进行偶联反应；而表面为羟基的载体，则可采用 N，N - 二乙氧基氨丙基三乙氧基硅烷处理而活化。采用类似的方法，可以分别把载体表面含羟基，氨基，醛基，肼基，巯基等活性基团活化，然后再进行不同的键合反应，下面就几种常见的表面化学活化进行说明。

1.有机胺活化芯片。带有羟基（ - OH）的硅玻片表面首先用 PDMS 或 N，N - 二乙氧基氨丙基三乙氧基硅烷处理，转化玻片表面的羟基基团成为活化的氨基基团，另外也可用酸溶液来处理表面氨化的玻片，使其带正电荷而活化，使得 DNA 分子能通过静电的方式而与芯片表面相连，并且这种连接在紫外线的作用下能转化成共价连接及电荷相互作用。其过程见下图。

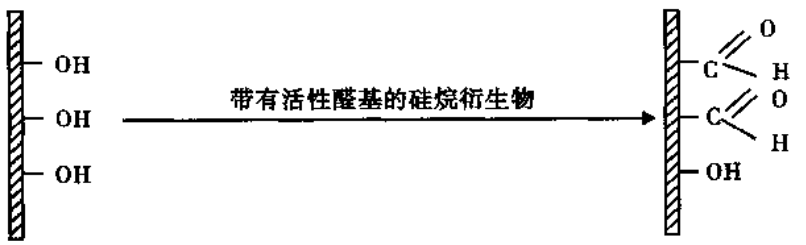
（1）活化：

(2)与靶分子键合：

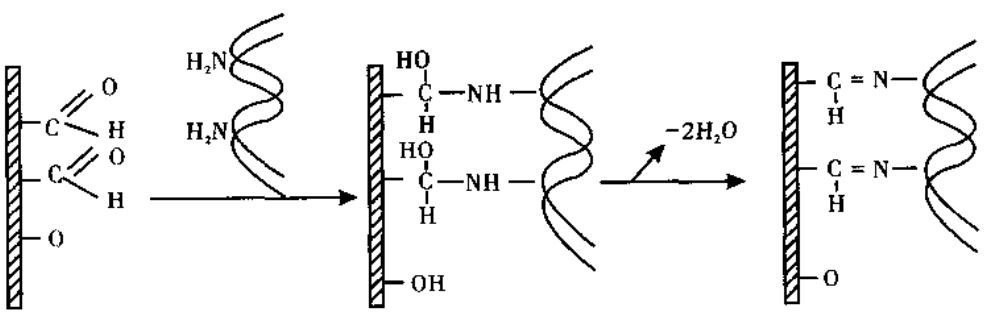


2.甲醛(Formaldehyde)活化的芯片。带有羟基的玻片也可首先被带有甲醛基的硅烷衍生物活化，表面带有醛基的玻片能与 DNA 或蛋白质的氨基相互作用而形成共价连接，然后通过脱水作用而形成 Schiff 碱的共价连接，使靶分子牢固地结合在玻片上。其过程见下图。

(1)活化：



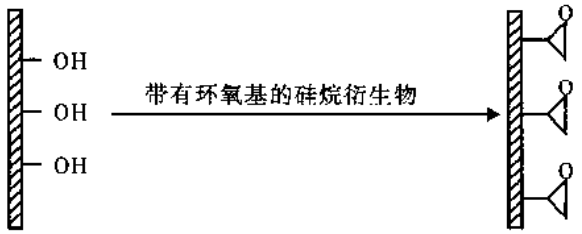
(2)与靶分子键合：



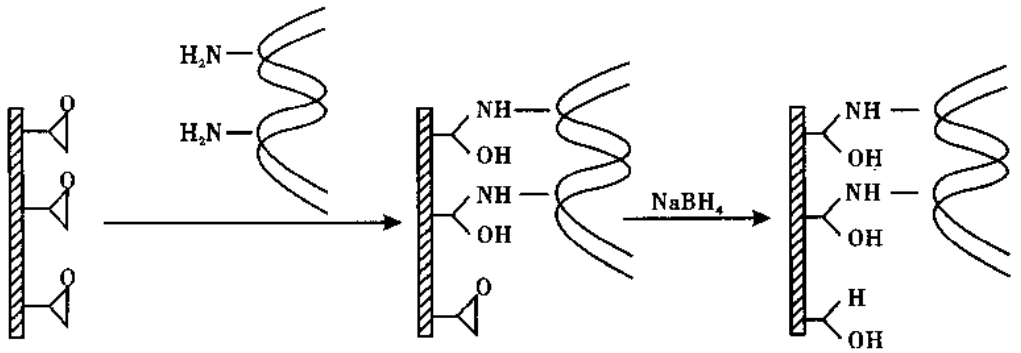
3.活化的表面带有环氧基(epoxy)树脂的芯片。通过与带有环氧基团的硅烷反应，将环氧基基团引入玻片的表面，也是目前常用的方法。带有氨基的 DNA 或蛋白质分子通过进攻环氧基基团而形成氨基键的连接，从而使分子牢固地固定在玻片的表面，为了消去剩余的环氧基

团，用硼氢化钠还原可使之变成羟基基团。其过程可表示如下。

(1)活化：

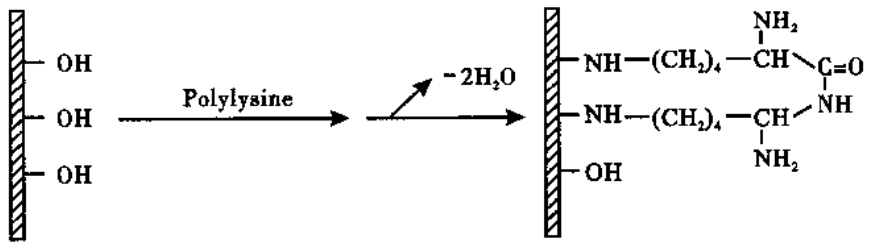


(2)与靶分子键合：

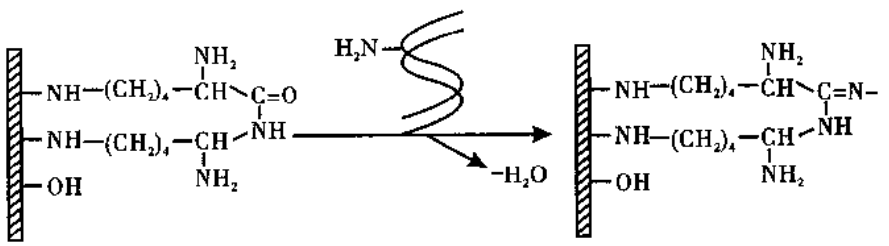


4.活化的表面带有赖氨酸基团的芯片。这是目前最常见的芯片活化方式之一，与有机胺活化的芯片相似，其原理是先将玻片表面的羟基转化成带活性赖氨酸基团的表面，然后 DNA 分子上的氨基基团与芯片上的肽键相互作用而形成共价连接，其过程如下：

(1)活化：

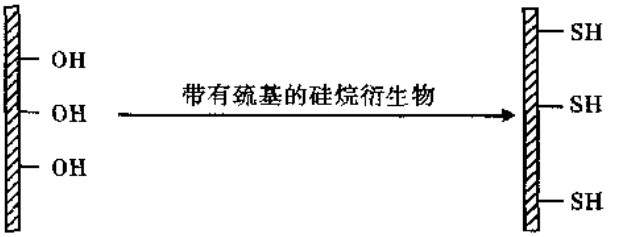


(2)与靶分子键合

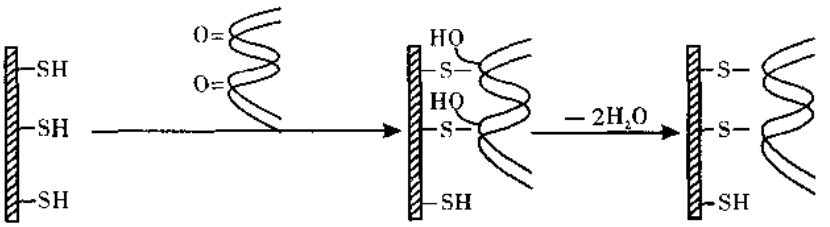


5.活化的表面带有巯基基团的芯片。带有羟基(－OH)的硅玻片表面也可以首先用带有－SH 的活性硅烷处理，转化玻片表面的羟基基团成为活化的巯基基团，DNA 分子与－SH 基团形成共价连接面固定靶分子。其过程见下图。

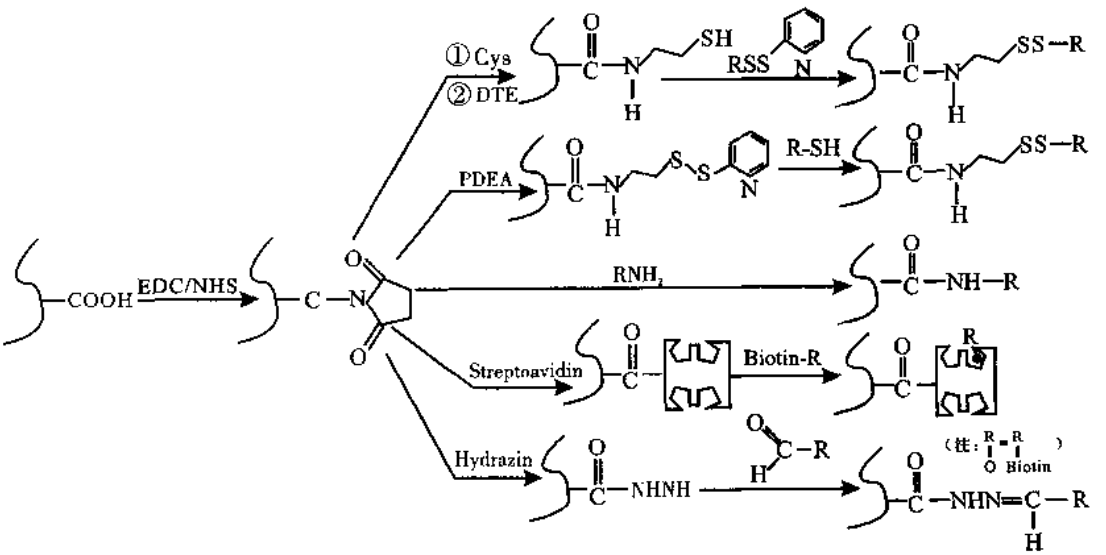
(1)活化：



(2)与靶分子键合：



6.其他。如表面为羧基的载体，则可通过下列反应进行活化：



在上述化学修饰中提到的硅烷化试剂，又称硅烷偶合剂，是分子中具有两种以上不同反应基团的有机硅单体，能起到把有机材料和无机材料进行化学偶合的媒介作用。一般具有RSiX3的化学结构。X表示水解基团，如甲氧基、乙氧基等，能与无机材料(玻璃、金属、二氧化硅)进行化学结合。R为有机官能团如氨基、羟基、巯基等，能与有机物质结合。代表性的硅烷化试剂有：

氨丙基三甲氧基硅烷　　$H_2NC_3H_6Si(OCH_3)_3$

氨丙基三乙氧基硅烷　　$H_2NC_3H_6Si(OC_2H_5)_3$

巯丙基三甲氧基硅烷　　$HSC_3H_6Si(OCH_3)_3$

N－氨乙基氨丙基三甲氧基硅烷　　$H_2NC_2H_4NHC_3H_6Si(OCH_3)_3$

三、载体修饰后的固定及选择原理

尽管芯片载体修饰有很多方法，最常用的有两种仍属醛基或氨基修饰的玻璃片载体。

醛基修饰的玻片主要利用 Schiff 碱结合的原理。在中性室温条件下，玻璃片上的醛基与靶分子的氨基反应生成弱 Schiff 碱而将待加样的分子固定下来。这种玻片可以制备 cDNA 芯片，玻片表面的醛基可与 DNA 上的碱基 G、C、A 的芳香胺反应形成 Schiff 碱；还可以通过与氨基修饰的 DNA（或寡核苷酸）反应形成 Schiff 碱而得以固定。如果用醛基修饰的玻片制备寡核苷酸芯片，则必须先在寡核苷酸的 5′端连接氨基碳臂后才能加样固定。

氨基修饰的玻片多用于制备 cDNA 芯片，氨基在中性条件下带正电荷与 DNA 分子中带负电荷的磷酸基团结合形成离子键而被固定，经 UV 交联后进一步提高固定率。

多聚赖氨酸修饰的玻片主要通过物理吸附固定 DNA 分子。多聚赖氨酸富集正电荷，易于吸附带负电荷的核酸片段，适于固定几百 bp 的 cDNA 或 PCR 产物，在紫外照射的能量作用下，多聚赖氨酸可以和 DNA 上的胸腺嘧啶碱基之间形成共价键进一步提高固定率。但该类芯片不适于固定较短的寡核苷酸芯片。国内有人将多聚赖氨酸修饰的玻片与氨基修饰的玻片对 cDNA 固定率和检测灵敏度进行比较分析，证明后者对 cDNA 的固定率要优于前者。

巯基化修饰的玻片需要先对玻璃基片进行洁净处理如酸处理、氨水处理等，然后用 MPTS 缓冲液对玻璃基片硅烷化处理，使其表面上形成巯基，而后再与相应的生物活性分子进行键合反应。

此外在载体表面还可以通过分子自组装构建功能化的单分子层，以保证芯片的均质性。

有很多经过表面处理的商品化的载体可供选用：TeleChem ArrayIt Super Microarray Substratcs（25 mm × 76 mm）采用高清洁度，超平表面的玻片并进行化学修饰，各项技术指标居同类产品前列。SuperClean（Cat. No. SMC－25，etc.），SuperAmine（Cat. No. SMM－25，etc.）和 SuperAldehyd（Cat. No. SMA－25，etc.）三种规格可偶联核酸、蛋白、甚至细胞，且偶联过程可以在室温及中性条件下进行。

Clontech 提供的 DNA－Ready Type I Slides（Cat. No. 7880－1，5 slides/box）是氨基修饰的，能结合 200 bp～10 kb 的修饰和未经修饰的 DNA，DNA－Ready Type II Slides（Cat. No. 7881－1，5 slides/box）是特殊氨基修饰的，能结合 50 bp～10 kb 的修饰和未经修饰的 DNA。

另外 CEL 公司也提供醛基化修饰的玻片 Silylated Slides（Cat. No. CSS－25）。

第二节　生物芯片制作方式

要将样品加到所选择的修饰好的芯片载体上的方法主要有两类，现分别以微阵列芯片和微流体芯片两大类为例介绍如下。

一、微阵列芯片及其制备技术

分子微阵列芯片是以生物技术为基础，在芯片表面固定一系列的可寻址生物分子阵列即探针阵列，可以是蛋白、糖、脂类和细胞等，通过与样品中的靶分子作用来获取生物信息。

通常微阵列制备方法包括离片合成法、原位合成法等，后者包括光引导原位合成法、压电打印原位合成法以及分子印章法；而离片合成法又包括点接触法及喷墨法。

(一)离片合成法(合成后加样)

离片合成法是将预先制备好的核酸、多肽、蛋白等生物大分子再用特殊的自动化微量点样装置将其以较高密度互不干扰地印点于经过特殊处理的玻片、尼龙膜、硝酸纤维素膜上,并使其与支持物牢固结合。支持物需预先经过特殊处理,例如多聚赖氨酸或氨基硅烷处理等;亦可用其他共价结合的方法将这些生物大分子牢牢地附着于支持物上。

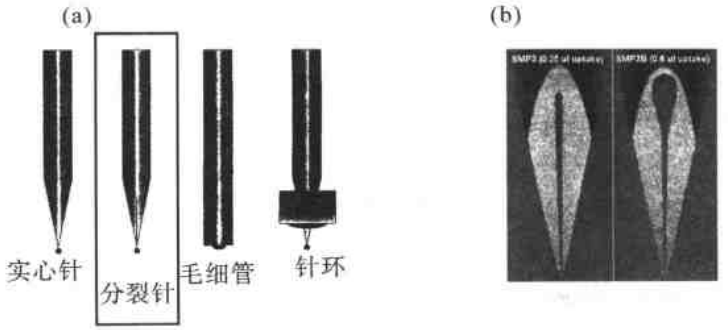


图 15-1 接触式点样方式

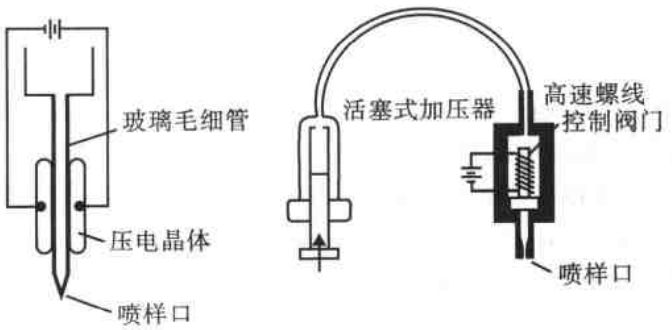


图 15-2 常见的两种非接触式点样方式

从图 15-1,15-2 中可见点样的方式分两种,其一为接触式点样,即点样针直接与固相支持物表面接触,将 DNA 样品留在固相支持物上;其二为非接触式点样,即喷点,是以压电原理将 DNA 样品通过毛细管直接喷至固相支持物表面。有效地将 DNA 固定在固相支持物上对 DNA 芯片的制作是很关键的。目前报道的固定方法有物理吸附、共聚合、络合以及共价结合,其中共价结合最为常用。在共价结合固定中,只是在很少的情况下生物分子直接偶联在固相支持物上(如巯基修饰的 DNA 固定在金表面);而通常是先在载体表面修饰一个功能化的单分子层,生物分子通过化学吸附固定在单分子层上。目前有两种系统用于共价连接固定:在羟基化表面的烷基硅氧烷单层和在金表面的烷基硫醇单层。离片合成法制作芯片的工艺比较简单便于掌握、分析设备易于获取,适宜用户按照自己的需要灵活机动地设计微点阵,用于科研和实践工作。

(二)原位合成法

1. 光引导原位合成法。此法是照相平板印刷技术(photolithography)与传统的核酸、多肽固

相合成技术相结合的产物。可用于 DNA 芯片的合成，也可以用于肽芯片的合成，整个合成均在芯片表面完成。基片在聚合反应前要先使其表面衍生出羟基或氨基（视所要固定的分子为核酸或寡肽而定）并与保护基团建立共价连接。光线通过一个掩膜，由于掩膜具有间隔的不透明区和半透明区，当光线通过掩膜上的半透明区时，基片上的光不稳定保护基团光解脱落，被保护的氨基酸或短肽、核苷酸等暴露于基片表面，加入反应底物，合成反应就会发生。不断地利用不同的掩膜并重复上述步骤，多肽或寡核苷酸序列就进行了在芯片原位合成的过程。

Affymetrix 公司率先开发寡聚核苷酸原位光刻合成技术，是生产高密度寡核苷酸基因芯片的核心关键技术。技术原理是首先使支持物羟基化，并用光敏保护基团将其保护起来。每次选取适当的蔽光膜（mask）使需要聚合的部位透光，其他部位不透光。这样，光通过蔽光膜照射到支持物上，受光部位的羟基脱保护而活化。因为合成所用的碱基单体分子一端按传统固相合成方法活化，另一端的 5′羟基则受光敏保护基团的保护，所以发生偶联的部位反应后仍旧带有光敏保护基团。因此，每次通过控制蔽光膜的图案（透光与不透光）决定哪些区域应被活化，以及所用单体的种类和反应次序就可以实现在待定位点合成大量预定序列寡聚体的目的。其基本原理如图 15－3 所示

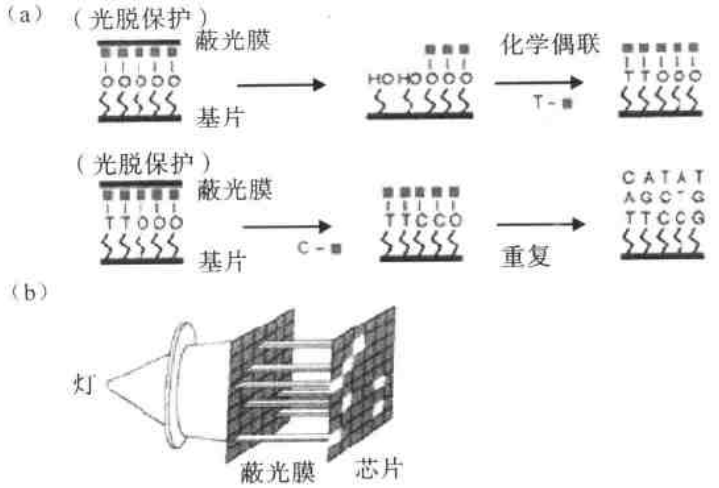


图 15－3　寡聚核苷酸原位光刻合成示意图

该方法的主要优点是可以用很少的步骤合成极其大量的探针阵列。如合成 65 536 个探针的 8 聚体寡核苷酸序列仅需 4×8＝32 步操作，8 h 就可以完成。不过，该方法每步合成反应产率比较低，不到 95%。因此探针的长度受到了限制，只能合成 30nt 左右长度的寡核苷酸。在此基础上，有人将光引导合成技术与半导体工业所用的光敏抗蚀技术相结合，以酸作为去保护剂，将每步合成产率提高到 99%，但制造工艺复杂程度增加了许多。所以如何简便地提高合成产率是光引导原位合成技术有待解决的问题。

2.压电打印原位合成法。其方式类似于喷墨打印机，合成原理与传统的核酸或寡肽固相合成技术相同。四个墨粉鼓分别带有四种不同的核苷酸（A、T、C 和 G），也可以是染色体 DNA、抗体和 cDNA 等。打印喷头将合成用的四种核苷酸按预定程序喷印到活化的基片上特定位点。喷印上去的试剂即以固相合成原理与该处支持物发生偶联反应，然后用酸去保护，重复此操作可以在特定位点合成出可长达 40～50 nt 的探针。其过程如下图 15－4 所示。

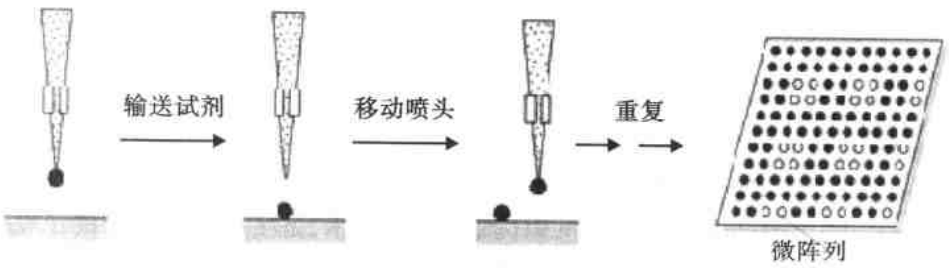


图 15－4　压电打印原位合成法工作方式

3．分子印章原位合成法是中国东南大学吴健雄实验室发明的，分子印章法工作原理见示意图 15－5，其方法是根据所要制备的微阵列，主要是寡核苷酸阵列，设计出一组不同的图形的微印章，然后按预先的设计把不同的化学单体（如含不同碱基的核苷酸单体），涂在不同的印章表面，然后依次压印在基板表面。在催化剂的作用下每次压印是一次化学联接反应。每次压印后，基板表面的化学基团应重新活化，以便进行下一次压印，在基板表面的分子活性基团上连接新的化学基团。通过一系列可控的压印过程，可以在基片上的指定位置合成所需的化合物阵列微单元，最终得到所需的寡核苷酸微探针阵列。

上述 3 种原位合成技术所依据的固相合成原理相似，只是在合成前体试剂定位方面采取了不同的解决办法，并由此导致了许多细节上的差异。但这 3 种方法都有在不同聚合反应之间如何精确定位及如何对合成的探针进行质量评估的问题，同时原位合成每步合成产率有限。但原位合成的短核酸探针阵列具有高密度，杂交速度快、效率高、较好的阵列批间连续性等优点，常用来进行突变检测、多态性分析、表达谱检测、杂交测序等需要大量探针和高的杂交严谨性的实验。

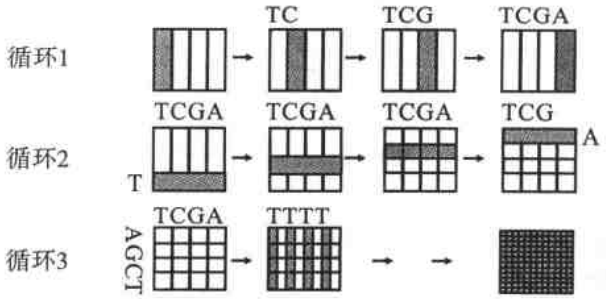


图 15－5　分子印章法原理示意图

上面所介绍的离片合成法与原位合成法是目前作为主要的两大类芯片制作方法，各具特点，其中原位合成法因其技术难度大、成本投入高等原因，只有少数大公司采用，但具有较高的精度和密度，可直接用 cDNA 数据库中的信息合成寡核苷酸，在 1.6 cm^2 的面积上可以合成 40 万个样品。离片合成后点样，虽然点样密度较低，每平方厘米可含 2 500 个样品，但由于其成本低，操作方便，一般的公司及研制单位广泛采用的点样方式仍是多种多样的离片合成法。表 15－1 对它们的特点作了较详细的比较。

表 15－1　离片合成法与原位合成法的比较

项　目	离片合成法	原位合成法		
		光引导原位合成法	压电打印原位合成法	分子印章法
芯片密度	中低密度	高密度	高密度	高密度
探针种类	寡核苷酸、cDNA、蛋白质	寡核苷酸 短肽	寡核苷酸、抗体、短肽、cDNA	寡核苷酸 短肽
探针长度	不限	约 25 nt	不限	较短
能否对探针评估	能	不能	不能	不能
工艺复杂性	简单	复杂	较复杂	较复杂
应用	基因表达 突变检测	基因表达 突变检测 测序	基因表达 突变检测	基因表达 突变检测

二、微流控芯片及其制备技术

(一)概介

微流控芯片以分析化学和分析生物化学为基础，应用微电子加工技术，在微芯片上加工出毫米级甚至微米级的容器、泵、阀、管道等结构网络，将样品制备、生化反应以及检测分析的整个过程集约化形成微型分析系统。现在已有由加热器、微泵、微阀、微流量控制器、微电极、电子化学和电子发光探测器等组成的芯片实验室问世，并出现了将生化反应、样品制备、检测和分析等部分集成的生物芯片实验室。

目前这类芯片往往是一个面积 2～3 cm^2 的细片，其材料可以是硅、玻璃、塑料等，通过蚀刻或印模的方法在细片上刻制切面不大于 50 μm 的通道等微结构，然后在芯片上结合盖板封闭。

(二)芯片的微加工技术

通过沉积、蚀刻、微铸和摹制等方法在基片表面或里面产生微型结构就是微刻技术。常用的方法是光刻平板印刷技术和软石板印刷术。

光刻平板印刷技术由光敏胶涂抹、光刻、刻蚀 3 个基本步骤组成，先将光敏感聚合物均匀涂抹于硅片或玻璃上，溶剂蒸干后，形成一层光敏层，在光敏层的表面上覆盖掩膜，在 UV 的照射下，可以选择性地将光敏层刻成相应的图案，再用化学刻蚀的方法可在硅片上刻制相应的凹槽，完成后除去光敏层即可。

软石板印刷术包括所有用于微结构成型的非蚀刻技术，如印刻法和微接触印制法。印刻法是一种低成本大批量制作塑料微芯片的方法。其作法是先在硅片上刻制微通道阵列的凸版三维结构，然后在塑料上印制得到设计的微通道阵列。微接触印制法使用平板橡胶印模将化学式样刻制到基片表面，相比印刻法，这种方法刻得到更高精度的微刻结构。

盖片通过阳极键合或热结合固定在刻有微刻结构的基片上，形成封闭的三维结构。阳极键合法是利用静电作用将玻片和硅片或金属粘合在一起，使得两者之间形成共价结合，而热结合法一般使利用粘结剂来完成。

通过上述技术，现在可以在厘米尺寸的芯片上构建诸如色谱或电泳分离柱、PCR 反应室、泵、阀、过滤器等。

(三)微流控芯片实例

目前研究较多的此类芯片有生物电子芯片、药物控释芯片、毛细管电泳芯片和微缩芯片实验室等。生物电子芯片是通过微电子加工技术,在热氧化的硅基片上构建电极阵列,并在每个电极上通过氧化硅沉积和蚀刻制备出样品池。将连接链霉亲和素的琼脂糖覆盖在电极上,在电场作用下生物素标记的探针即可转运到特定电极与目的片段杂交,即电子杂交。由于电场的作用使得 DNA 选择性传送,同时使杂交的速度、特异性更高。Cheng 等利用该芯片对病人全血分析,检测到 *E. coli*、溶壁微球菌和表皮葡萄球菌的 DNA。并在此基础上,创造性地将电场、磁场与芯片结合起来。此外生物电子芯片还常常用来对癌基因、地中海贫血病等遗传性疾病进行单碱基突变分析。

缩微芯片实验室是生物芯片工艺技术发展的最高水平。目前可用于细胞分离、样品预处理、血气检测、PCR 扩增和毛细管电泳的微流体芯片均有报道,但它们功能相对过于简单,只能作为缩微芯片实验室的雏形。今后的方向是将样品制备、生化反应以及检测分析的整个过程集约化形成微型分析系统,即缩微芯片实验室。微小的尺寸和各种技术的整合使得缩微芯片实验室具有很多显著优点,首先样品的要求量只需微升甚至纳升级、分析过程快速化、自动化、集成化,携带方便等等。但也有一些不足,由于只用了微量样品,因此提高了检测的难度,对污染也更加敏感等。

三、其他类型芯片制作方式

如通过导电的吡咯单体的聚合形成微阵列、或者将寡核苷酸固定在光纤纤维的一端,然后将光纤纤维捆绑成束,组成由不同寡核苷酸集合体的微阵列,但这种制备方式实质上还是属于合成后加样的模式。

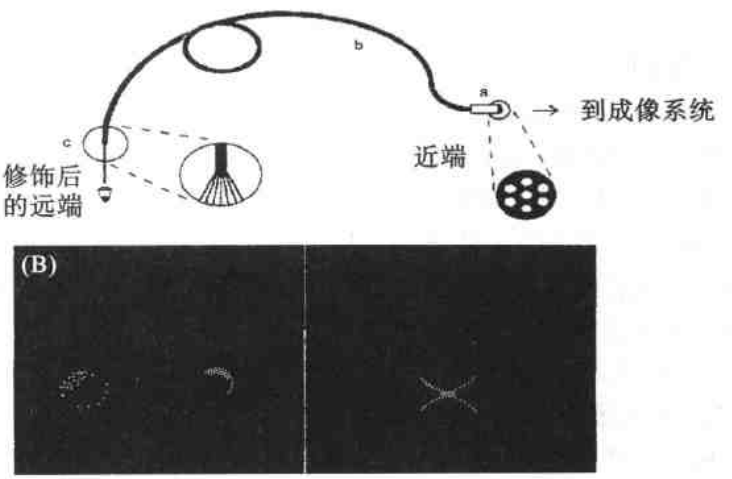


图 15-6 光纤传感器阵列系统的结构图

(A)光纤生物传感器图解。(a)纤维近端与成像系统纤维夹的不锈钢管相连;(b)光纤束长 3 英尺(ft)可用于远距离探测;(c)保护橡胶管从纤维束远端移开后,开始进行操作。

(B)CCD 相机对插入缓冲体系中的光纤传感器微阵列成像。(左)白色光线传导通过阵列远端成像;(右)缓冲液经 480 nm 波长激发后在 530 nm 背景荧光成像。光线通过覆盖层时产生的一些干扰,在离微阵列远端的接受部分不形成荧光信号。

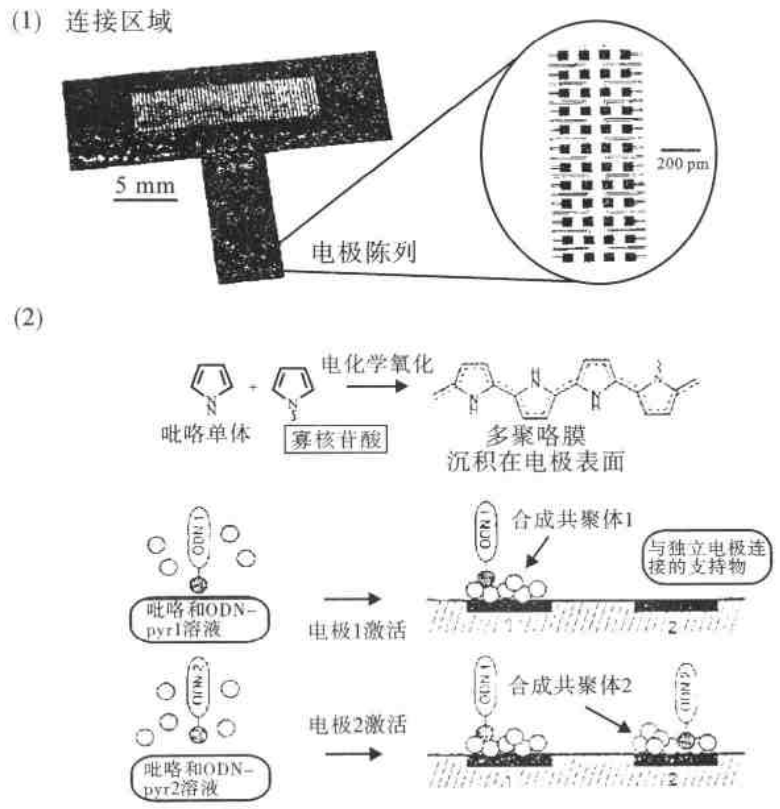


图 15－7　电化学

(1)含有 48 个金微电极阵列硅片的示意图;(2)用于肝炎 C 病毒基因分型的多聚吡咯 DNA 芯片

(A)吡咯聚合原理;(B)可寻址的寡核苷酸(ODN)电诱导性共聚合;首先 ODN－pyr1 在电极 1 共聚合,然后洗涤细胞,电极 2 激活后 OND－pyr2 进行聚合。

四、基因芯片和寡核苷酸芯片制备原理

微阵列基因芯片的制备包括待固定样品的制备与纯化、点样、点样后的固定处理(例如紫外交联)、封闭防止非特异性吸附等步骤,下面以医学研究中最常用的 cDNA 芯片和寡核苷酸芯片为例进行说明(更具体的内容将在下节讨论),对其他类型的芯片制备步骤仅以光纤 DNA 生物传感器微阵列的制备为例作一简介。一般可以购买商品化的已制作好的 cDNA 芯片或寡核苷酸芯片,如有特殊需要也可以自己制备。

(一)cDNA 芯片的制备流程

1.用 PCR 扩增 cDNA 克隆(扩增片段长范围在 800～3 000 bp 之间),反应完成后,通过 G－50 柱纯化除去未掺入的游离核苷酸以获得 cDNA 片段。

2.PCR 产物用异丙醇纯化后进行点样。选择适宜的芯片载体,再进行硅烷化处理,得到表面氨基修饰的芯片。样品 cDNA 可溶于水中,也可以溶于 3×SSC 的盐溶液中。根据有些作者的经验,虽然用水溶解的样品在玻片上的固定率高,但是均匀度、整齐度及面积大小均不如溶解在盐溶液中,建议还是使用盐溶液。点样液的浓度一般为 0.5 μg /μL。点样可根据不同条件,采用不同的点样仪,手动或自动点样。

3.点样后处理。将点样后的玻片进行扫描,观察均匀度,然后经过水合、干燥等步骤处理。一般先水合 1～2 h,再在室温下干燥 30 min。后置于紫外交联仪中进行交联(60 mJ),再用 0.2%SDS 在室温下洗涤 5 min,最后用 ddH_2O 冲洗 2 次,每次 1 min,再进行扫描。或在水洗后再在 0.2%硼氢化钠溶液中处理 10 min,晾干备用。

点样后,也可根据实际情况,自己摸索最佳的洗涤条件。

4.芯片固定率的测定。为了评价 DNA 的固定效果,必须检测 DNA 的固定率。固定率特指加样后的芯片经洗脱后,与玻片结合的 DNA 总量占总点样量的百分比。

固定率 = (洗脱后的荧光值 / 洗脱前的荧光值)× 100%

为此,可将用 Cy3 和 Cy5 标记的一定长度的 DNA 样品,将 PCR 产物纯化后点于玻片上,每个样品点 20 个以上,用扫描仪相应的软件测定和分析不同条件下样品在洗脱前后的荧光强度,计算固定率。

5.用标记探针杂交实验进一步验证芯片的质量。根据上面实验所确定的优化条件,固定 cDNA 样品后,与人胎盘 mRNA 进行杂交实验,观察表达谱芯片的的信噪比,选择比值最高的进行实验。到此为止,一张可用于检测的 cDNA 芯片便制备成功。当然,如果进行大规模的制备,那芯片制备中的批间误差必须控制在一定精度内以控制质量。

(二)寡核苷酸芯片制备流程简介

1.芯片准备及点样。根据不同的需要,所合成的寡核苷酸探针的 5′端必须带有氨基修饰后才能点样。寡核苷酸芯片的制备主要利用 5′端带氨基的寡核苷酸通过接头分子与氨基修饰玻片共价键连接,接头分子(linker)可以是 1,4－苯二异硫氰酸酯、戊二醛或丁二酸酐等。

如果将氨基修饰的玻片浸泡于 5%的戊二醛中,在室温下放置 30min,水洗二次,室温干燥后就得到醛基修饰的玻片,可进行点样。但是处理时戊二醛的两个醛基有可能同时与玻片上的氨基结合,而影响其中一个醛基与寡核苷酸探针的结合而减少探针固定容量,但是该法的优点是处理方便和价格便宜。

所需准备的样品浓度为 10 μM 溶于 50%DMSO 点样液中。

2.点样后玻片处理。60 ℃烘箱烤 4 h(封闭,并放卫生纸加少许水以保持湿度)后,玻片转移至硼酸钠溶液中(1.0 g 硼酸钠溶于 300 mL PBS,加入 100 mL 100%乙醇,用前临时配置),在摇床上缓慢转动处理 3 min,以减少游离醛基。再以 0.2%SDS 洗涤 2 min、以 dH_2O 洗涤两次,每次两分钟。离心甩干,储存于暗处,并保持干燥。

(三)光纤 DNA 生物传感器微阵列的制备流程简介

光纤生物传感器微阵列(fiber－optic biosensor microarray)也可以同时分析大量 DNA 序列。如果将每一根光纤纤维的远端都分别固定上不同的寡核苷酸探针,再将这些修饰后的光纤纤维捆绑成束,构成寡核苷酸的微阵列。荧光标记的互补寡核苷酸与微阵列杂交,并通过 CCD 照相机监视荧光强度的变化。这种类型的微阵列检测同样快速灵敏,可在不到 10 min 的时间内同时检测出 10 nmol/L 微量的众多 DNA 序列。

在制备修饰光纤纤维时,先将纤维捆的两端抛光,洗净,远端放入 10%(氨基)丙基三乙氧硅烷的丙酮溶液(v/v)中硅烷化 2 h,取出在丙酮洗涤几次后于空气中干燥 30 min。将纤维放入 1.25%的戊二醛－磷酸盐缓冲液(pH6.8,0.02 mol/L)中 30 min,然后再在空气中干燥 1 h,洗净后放于蒸馏水中,直至需要时再取出。

10 nmol 的 5′氨基修饰的寡核苷酸溶于 90 μL 的 0.1 mol/L 的硼酸钠缓冲液中(SBB)。加

入 5 nmol 的 cyanuric chloride(溶于 10 μL 的乙腈中),室温下反应 1 h,未反应的 cyanuric chloride 可通过三次离心超滤去除(3 000 dalton 的分子量截留值,Microcon 3, Amicon 公司),然后用 200 μL 的 0.1 mol/L 的硼酸钠缓冲液中(SBB)稀释,所得到的活化引物回收在 50 μL 的 0.1 mol/L 的硼酸钠缓冲液中(SBB),于 4 ℃下可放置 1 个月。

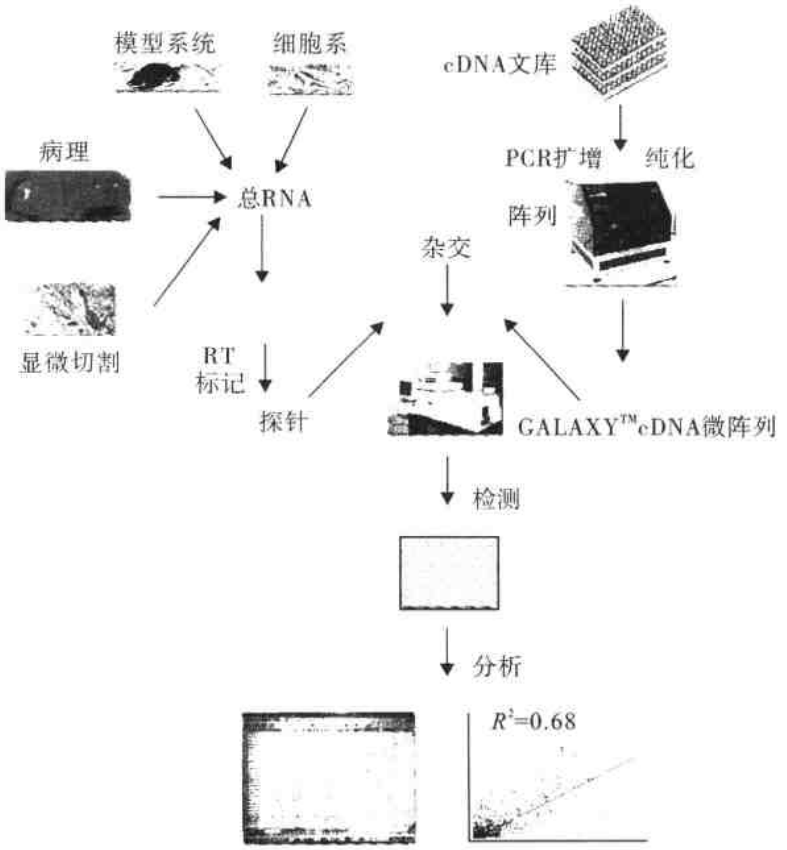
将待修饰的光纤纤维的远端浸泡在 5 个 2 mL 的 0.1 mol/L 的 SBB 中进行洗涤。然后将顶端浸入 10 μL 含 150 nmol 的被 cyanuric chloride 活化的寡核苷酸 - 0.1 mol/L 的 SBB 溶液中,1 ~ 2 h,取出后再转移至含 200 μL 90% DMSO、10% 1 mol/LSBB(pH8.3)、0.1 mol/L succinic anhydride 的溶液中,室温下反应 1 h。取出后光纤纤维的顶端将按照上面所介绍的方法在更换的 0.1 mol/L 的 SBB 中洗涤 2 次,接着在每一次都更新的含 0.1 mol/L NaCl 及 0.1% SDS 的 TE 缓冲液(10 mmol/L 的 Tris - HCl,pH 8.3,1 mmol/L EDTA)中洗涤 5 次,然后储于该洗涤液中直到使用前。于是功能化的寡核苷酸光纤顶端微阵列便可捆绑而成。

从上面所介绍的内容可见,不管是什么类型的芯片微阵列,其载体修饰方式及微阵列制作方式的基本原理及步骤都是相通的。

第三节　生物芯片检测的实验流程

对于表达谱芯片其检测大致流程包括(见图 15 - 8a):以下将从杂交探针的制备、样品制备、分子杂交等 3 个方面对芯片检测的实验流程作以介绍(图 15 - 8b)。

(a)



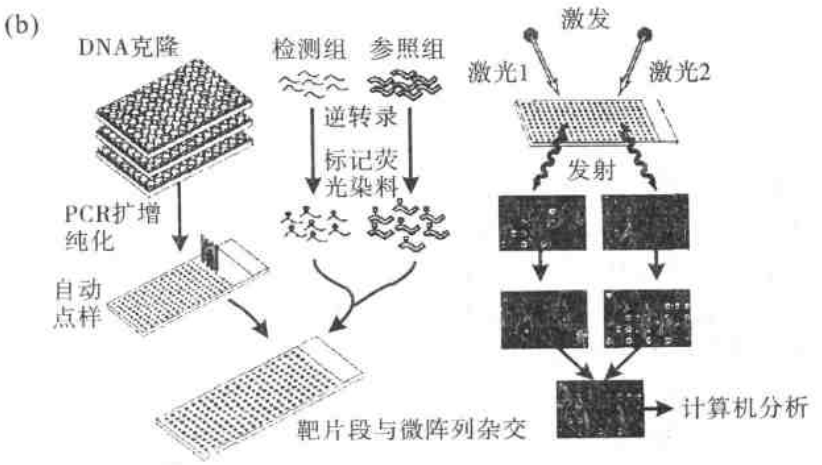


图 15-8 流程图

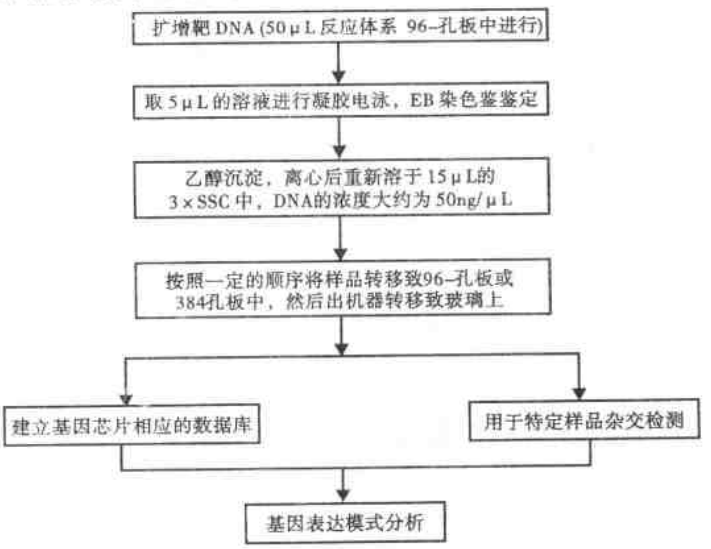
一、芯片样品的制备

获取大量的数据和有用的信息以及相对低的成本花费是在芯片制作上首要考虑的两点。样品的精细度和一致性往往确定了芯片的质量，不同的样品需根据其特性采用不同的方法来制备。下面就几种主要的样品制备进行讨论，为便于对各类芯片的全面介绍，我们这里将寡核苷酸芯片与 cDNA，基因组 DNA 芯片归为一类来介绍。

(一)DNA 芯片样品的制备

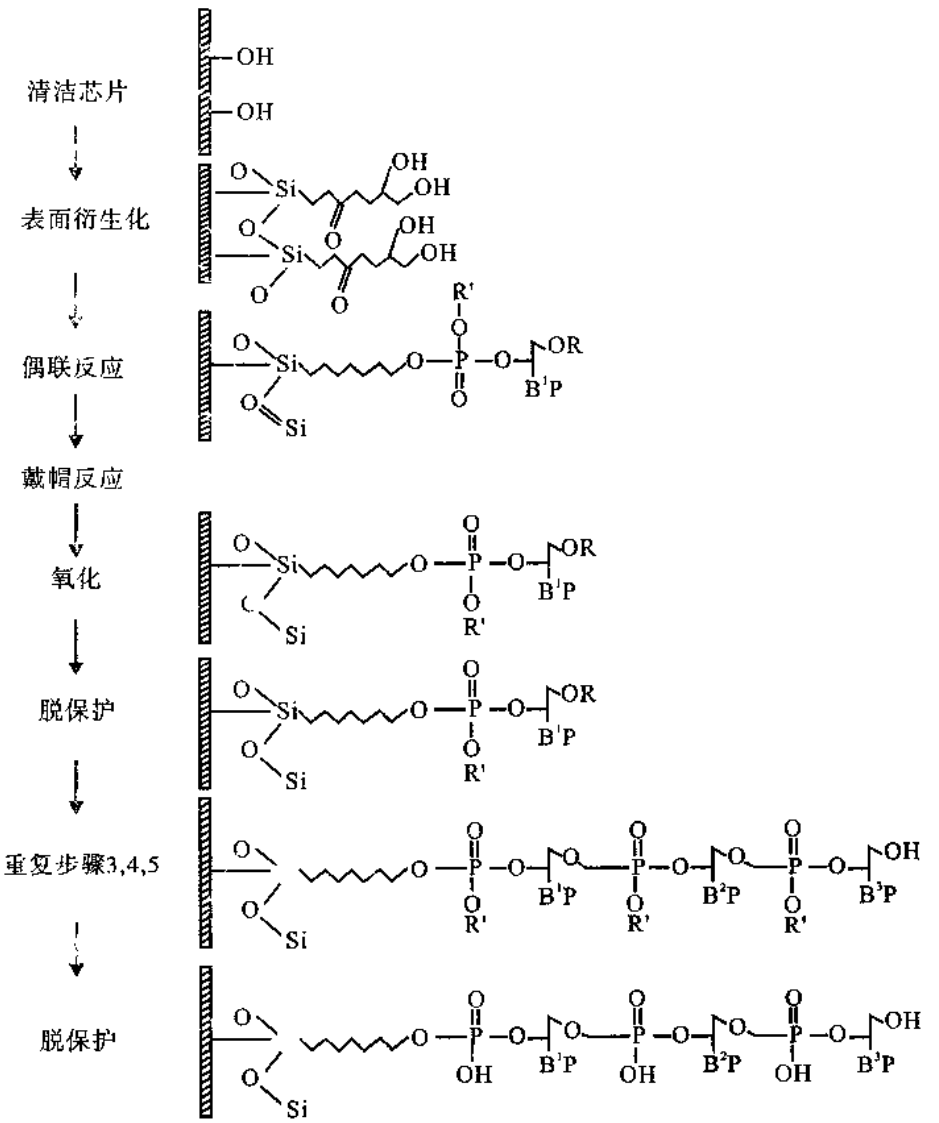
根据来源 DNA 样品主要包括 cDNA 样品，基因组 DNA 样品，以及合成的寡核苷酸 DNA 样品。

1. cDNA 样品的主要来源是：①已知基因的 cDNA 序列，通过搜集已知的基因或 ESTs(expression sequence tags)序列，然后 PCR 扩增(往往采用 96-孔板的方式进行 PCR 扩增)后用于芯片制作，这是目前芯片制作中的最重要方法，大部分的基因或 ESTs 均可从基因公司购得(如 research genetics inc.)；②从细胞中提取 mRNA 经逆转录后构建成 cDNA 文库，然后随机挑取克隆和进行大规模基因测序，根据测序结果来选择所需克隆进行芯片样品的制备，当然，直接用所构建的 cDNA 文库克隆进行样品的制作而不经测序也往往是研究者常采用的芯片样品的制作方法。如 cDNA 样品的制备可用图表示如下：

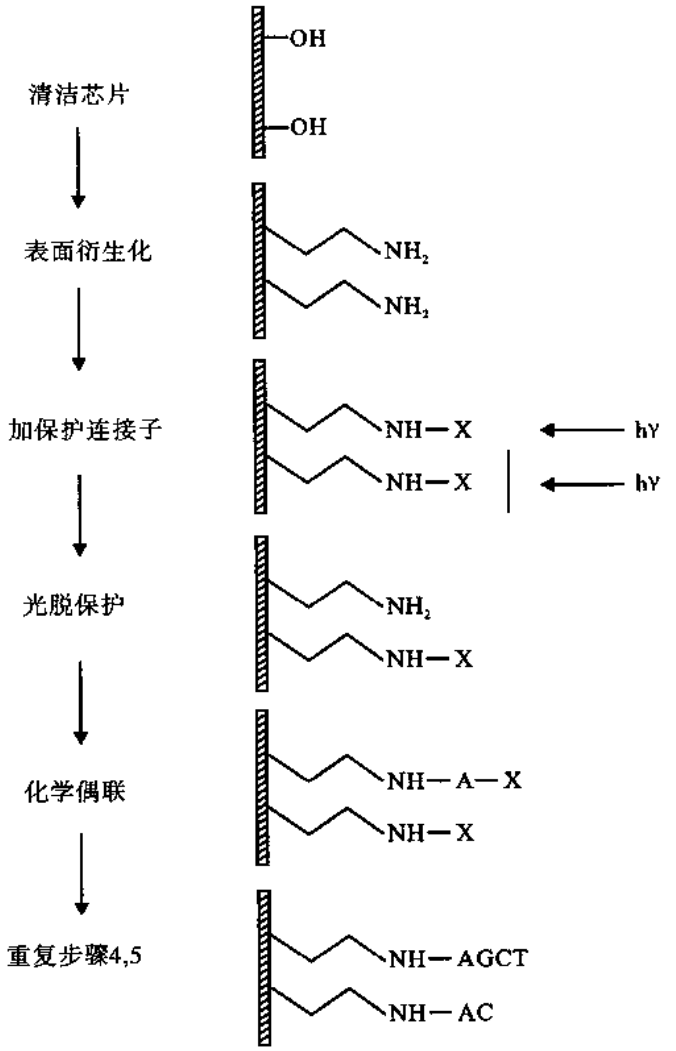


2.基因组 DNA 样品的制备则依据实验目的的不同而改变，如要检测 DNA－蛋白质相互作用或潜在启动子所在序列，可将所有的候选者序列以 PCR 的方法扩增成适合长度的片段，再将产物刻印至玻片上；如以检测点突变为目的，则可将被检测的基因组经 PCR 扩增各种不同的片断后转移至玻片上固定，也可利用寡核苷酸芯片的方法，直接在芯片上合成寡核苷酸，它包括热化学合成和光化学反应合成法，其过程见下：

(1)热化学法合成寡核苷酸芯片：



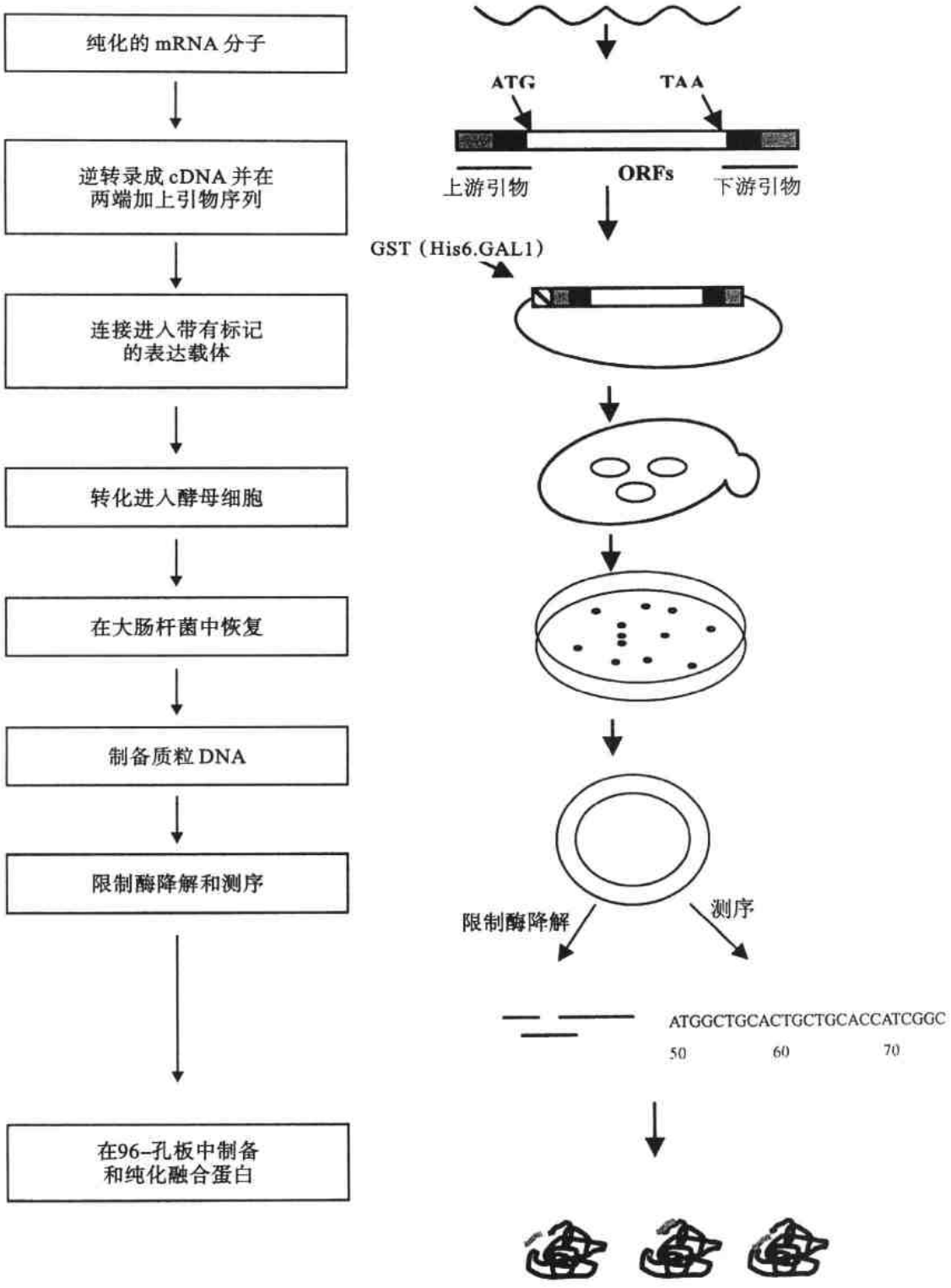
(2)光化学法合成寡核苷酸芯片：

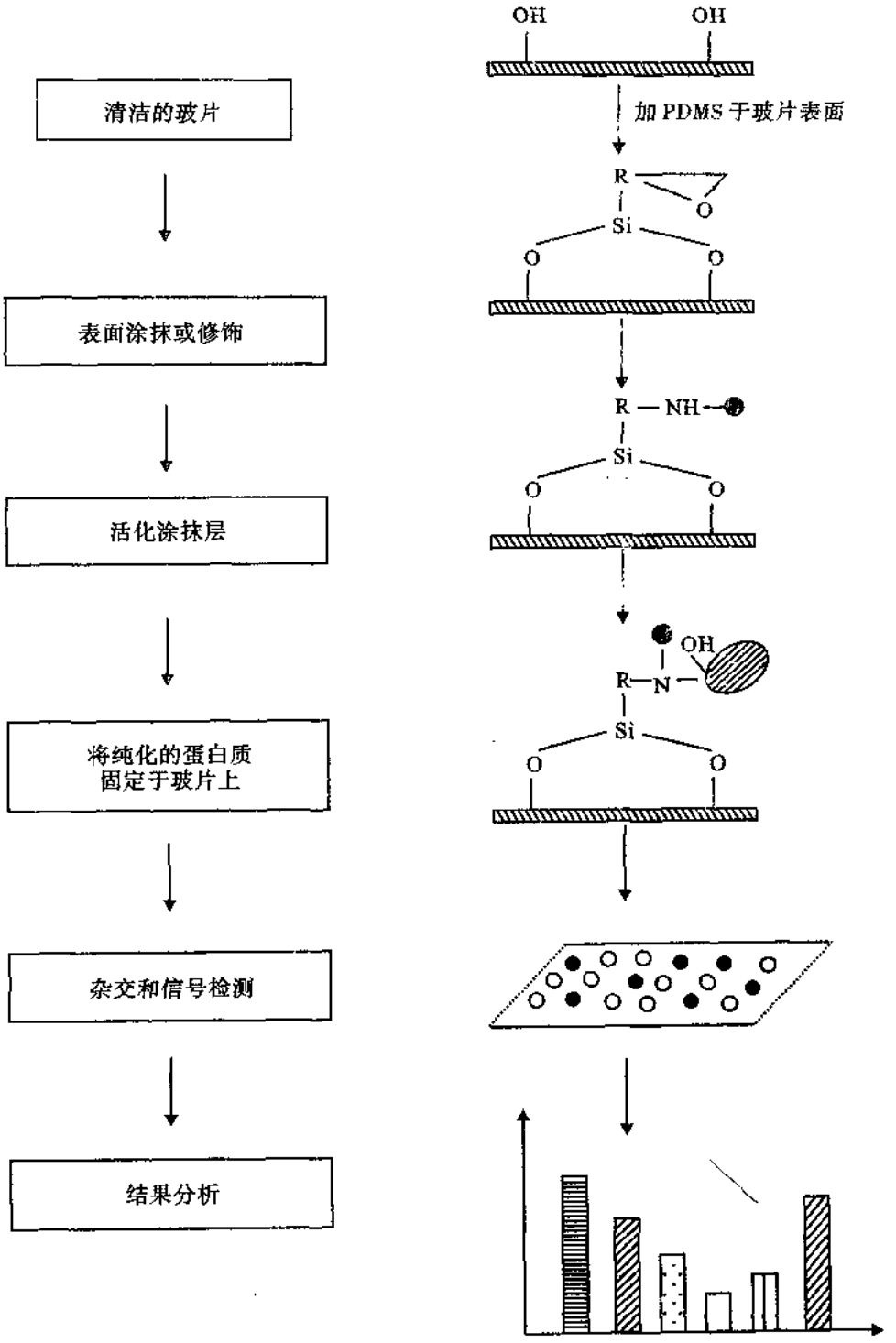


(二)蛋白质(或多肽)芯片样品的制备

虽然目前的技术允许方便的制造大量的基因表达克隆来进行蛋白质的生物医学功能分析,但到目前为止,在这方面的研究仍然很有限,下面仅以一些技术上成熟的例子来说明用于生物芯片蛋白质样品的制备。

当然,也可以在经修饰的玻片表面直接进行多肽的合成,这种方法在样品的制作上就要简单许多,只需一些经修饰的氨基酸即可,但在技术上较困难,需经一系列的特需的光定位反应,而且产量不高,在这里就不单独阐述。

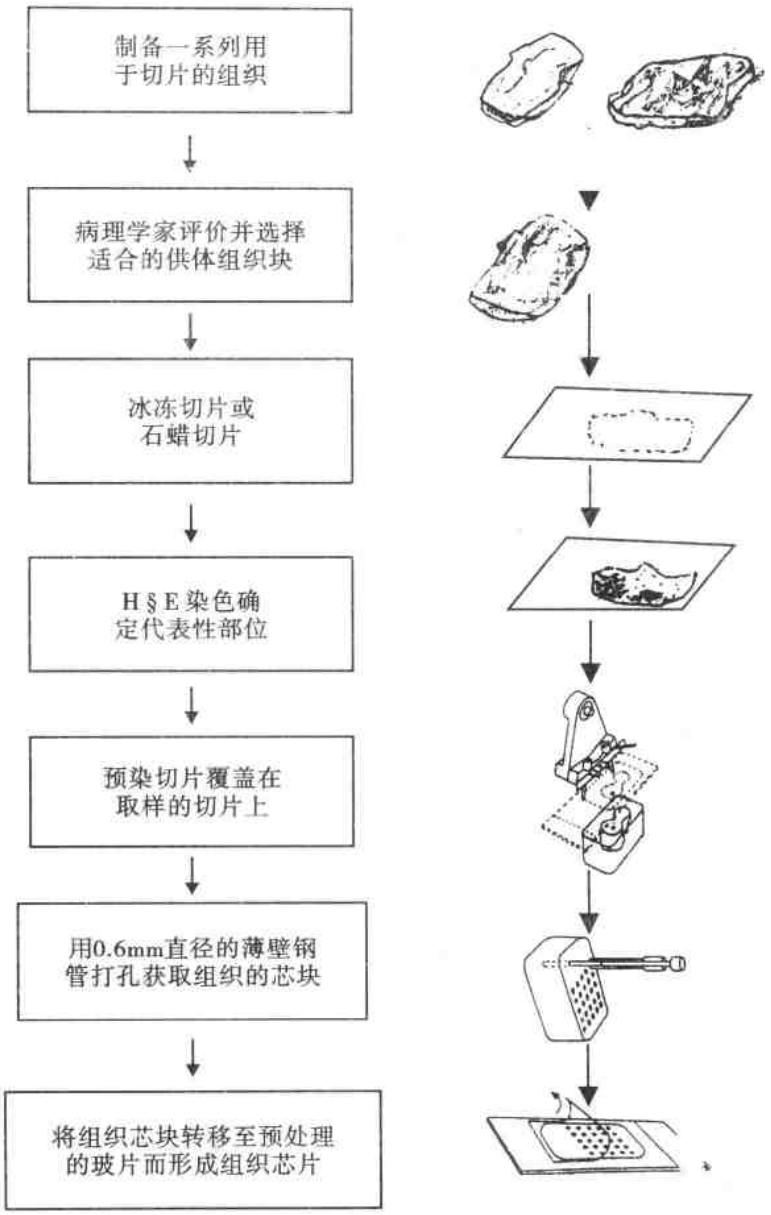

纯化的 mRNA 分子
逆转录成 cDNA 并在两端加上引物序列
连接进入带有标记的表达载体
转化进入酵母细胞
在大肠杆菌中恢复
制备质粒 DNA
限制酶降解和测序
在96-孔板中制备和纯化融合蛋白
ATG
TAA
上游引物
ORFs
下游引物
GST (His6.GAL1)
限制酶降解
测序
ATGGCTGCACTGCTGCACCATCGGC
50
60
70



(三)组织芯片(tissuemicroarray,TMA)样品的制备

组织芯片技术的发展是最近几年的事,与普通的组织切片相比,组织芯片的优点主要是:①通过改善组织资源的保护和利用,从而增加以组织为基础的分子水平的生物功能研究;②极

大地增加每次的实验样本的数目，从而使实验更为精确；③空间连续分布的组织芯片对于研究肿瘤的发生和演进，甚至在器官组织的发生上也具有极为广泛的应用前景；④国际互联网络技术为不同研究小组间进行诊断和资料共享提供了技术支持，而基于芯片的诊断分析结果将为异地诊断提供一个强有力的支持，但这项技术还处在探索阶段。制备组织芯片的样品与普通的组织切片类似，主要过程如下：



所形成的组织芯片可以用免疫组化，荧光原位杂交（FISH），RNA 原位杂交（RNA－ISH）等。应该说目前的组织芯片的样品制备过程 90% 的都是传统的病理学工作，其差别主要在于组织芯片可以包括大得多的数量和更高的精密度，当然一味的强调其差别是没有意义的。

二、杂交探针的种类和标记

基因芯片检测应用的是杂交原理，而核酸分子杂交则需要制备探针，直接用于检测靶基因或者检测外源病原体基因。为提高检测灵敏度，可采用两种办法，一是放大检测信号；二是增加模板量，这就需要扩增模板。下面重点介绍利用探针技术放大信号方面的研究情况。

（一）分支探针

一般而言，这类探针分子具有放大信号的功能，分子由两部分组成，即识别部分和报告部分，见下：

分支探针
- 识别靶序列
  - 靶序列短：杂交迅速
  - 靶序列长：完成杂交费时长，甚至过夜
- 携带报告分子
  - 携带越多，信号越强，灵敏度较高
  - 可用核素或非核素方法标记报告分子

根据上述原理可设计两类杂交法：①同时采用两种探针，一为捕捉探针，将其固定在载体上；另一探针的一半为识别序列，可识别并与靶分子结合；另一半则与捕捉探针互补结合。②直接将带识别靶序列的探针固定在载体上兼起捕捉作用。

探针除了有识别序列外，还需引入报告分子以供信号检测源检测。为此，可选用一个带有多官能基团的分支探针，每一个分支均与一个报告分子连接；分支探针的另一端则留出一个官能基团与载体上已固定的寡核苷酸探针（相当于上面提到的捕捉探针）相连。分支探针的使用使杂交信号得以成倍放大，见图 15－9(a、b)。

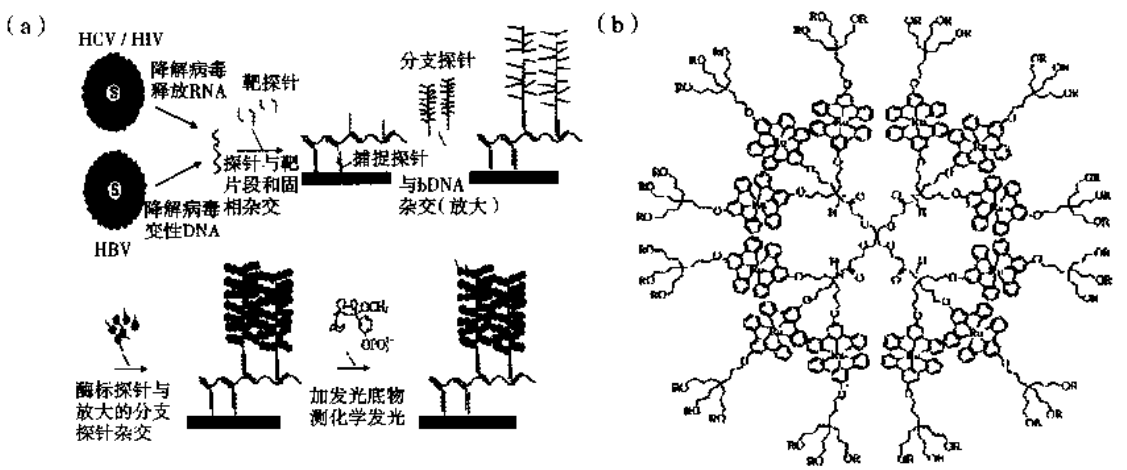


图 15－9　分支探针放大信号原理

(a)多探针夹心杂交原理；(b)含 12 个三联吡啶络钌的发光分子

（二）肽核酸（peptide nucleic acid）探针

从命名可见肽核酸是既像核酸也非核酸；既像多肽也非多肽的一类化合物，而是以中性酰胺键（肽键）为骨架的一类新的寡核苷酸的类似物，故称 PNA。PNA 与寡核苷酸的磷酸二酯键骨架不同的是它以甘氨酸结构单元为骨架；PNA 中的碱基部分则通过亚甲基羰基与主骨架相

连，碱基与骨架间隔 3 个化学键，相邻碱基之间相隔 6 个化学键(图 15－10)。

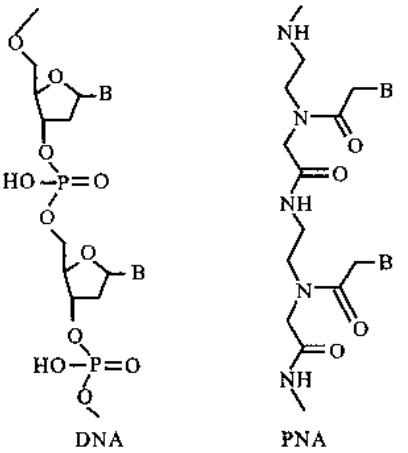


图 15－10　PNA 结构图

PNA 与天然核苷酸结构的比较如下：①与天然核苷酸结构相同之处是它们的碱基均能通过 Watson－Crick 氢键与互补 DNA 的特定核酸序列配对结合；②由于 PNA 分子不带电荷，互相之间及与核酸杂交时均不存在静电排斥作用，用它代替寡核苷酸探针固定在芯片载体上，对杂交反应具有更大的优越性。又由于 PNA 分子缺乏核酸分子所具有 5′→3′的极性，因此杂交时可以正向或/和反向方式与靶序列互补结合(以反向结合占优势)；③PNA 由于缺乏磷酸二酯键，能耐受核酸酶的降解，比较稳定；④PNA 也可以用固相肽合成技术合成，制备方便；⑤PNA 与靶序列杂交时，对碱基错配有很强的识别能力。例如一个 15 mers 的 PNA/DNA 异源双链中，有一碱基错配时，Tm 值将下降 8～20 ℃；(平均 15 ℃)，而对应的 DNA/DNA 复合物中，一个碱基错配时，Tm 值只下降 4～16 ℃(平均 11 ℃)。这种高度的区分能力可以建立以 PNA 为基础的分析点突变的方法；⑥PNA 能识别 DNA 双链上的特定碱基序列，形成局部的 PNA/DNA/DNA 或 PNA/DNA/PNA 三链复合物。

(三)分子信标

这是一种设计巧妙的新型荧光标记核酸探针。在长度为 15～30 mers 的序列特异性的寡核苷酸探针的 3′及 5′端，分别加上 5～8 mers 长度的互补序列，称为茎杆区。在自由状态时，由于茎杆区的存在，使探针分子形成发夹状结构，所以又称为发夹式探针；其形状又像海面上的灯塔，也有人称将这类探针称之为分子灯塔。

在分子信标的 5′及 3′末端分别连接荧光素分子(F)及猝灭剂分子(Q)。在分子处于自由态时，发夹结构的存在将两末端拉在一起，从而使(F)与(Q)靠近。由于分子共振和能量转移，使 F 和 Q 之间的能量传递交换几乎达到 100%，荧光被猝灭；当体系中由于靶分子的加入，分子信标中的环状区与靶分子的序列完全互补，形成异源双链杂交体时，信标茎杆区也随之被拉开，F 与 Q 间的距离增大，信标分子的荧光几乎 100%恢复。故分子信标在杂交前后的信号/背景比值差别极大，见图 15－11。

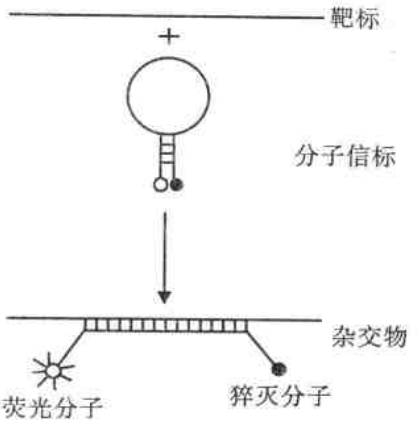


图 15－11　分子信标的工作原理

发夹结构的茎杆是由一段互补的短序形成的，当它与靶标杂交形成异源双链螺旋时就不能保持这种发夹结构。构型改变使茎杆臂拉开使猝灭分子离开荧光分子。

综上所述，信标分子一般包括 3 个部分：

1.识别部分。是与靶分子杂交的部分，长度一般为 15～30 mers。若在杂交双链中有一个错酸碱基，其热稳定性显著下降，故分子信标可用于检测点突变。

2.茎杆部分。位于信标分子两端，长度为 5～8 个序列互补的核苷酸。

3.荧光素、猝灭剂和连接臂。荧光素连接在信标分子的 5′端，常用 DABSY(二甲氨基偶氮苯甲酰基)荧光染料。荧光素、猝灭剂与核苷酸之间的连接，需要分别引入连接臂，如$(CH_2)_6-S-CH_2-O$及$(CH_2)_7-N-$两种类型的碳臂。

信标分子能在均相系统中实时测定出靶标的含量，而不必分离去除多余的信标分子。其主要应用如下：①实时定量 PCR 测定靶标浓度(图 15－12)；②PCR 扩增并进行点突变的检测；③多色荧光分子信标检测人的等位基因；④用分子信标 PCR 法目测测定结核杆菌的耐药性。

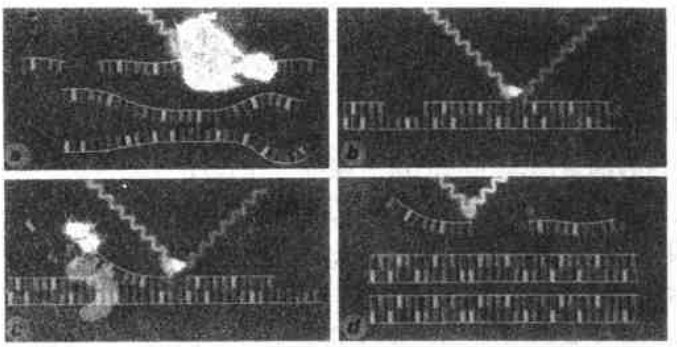


图 15－12　FRET 技术在定量 PCR 中的应用

(四)探针标记

现以常用的基因芯片为例进行说明。DNA 芯片检测的技术实际上就是核酸杂交技术，是由靶分子和探针组成杂交的基本杂交体系。因而在探针标记时，还需要经过体外扩增并加以标记才能在与芯片上的靶分子杂交后有足够的信号强度供检测用，特别是采用诊断芯片检测血液中微量病原微生物的基因组核酸时，杂交前一定要对样品进行扩增，并在扩增的过程中进行探针标记。探针的标记主要包括：

1. 标记分子为 RNA。采用逆转录过程将荧光素标记的核苷酸掺入到新合成的核酸 cDNA 探针分子中去。大致过程如下：①样品总 RNA 的提取：用 Gibco 公司的 Trizol 试剂一步法提取细胞或组织的总 RNA。详见说明书，主要步骤如下：细胞用冰冷的等渗 PBS(pH7.2)洗涤 2 次，加 Trizol 试剂裂解细胞，氯仿抽提一次，离心后吸出上层液相，异丙醇沉淀 RNA，75% 乙醇洗 2 次，空气中干燥，溶于缓冲液中，用紫外分光光度计测定 OD_{260nm}/OD_{280nm} 的比值，并用甲醛凝胶电泳观察 28 s、18 s、5 s 条带的比例，以确定总 RNA 的提取质量。②mRNA 的分离和纯化：用 mRNA 纯化试剂盒(例如 Quagen 公司的 Oligotex mRNA Midi Kit 或其他 Kit)，操作步骤见说明书。③ 逆转录生成 cDNA：参照 Schena 的方法逆转录生成 cDNA。在此过程中用 Cy3 - dUTP 标记对照细胞的 mRNA，用 Cy5 - dUTP 标记实验细胞的 mRNA。逆转录反应完成后，乙醇沉淀并溶于 20 μL 5 × SSC + 0.2% SDS 杂交液中。

2. 标记分子为 DNA。探针的标记物主要还是采用荧光素标记。标记的方法：①可以利用荧光标记的引物，进行 PCR 扩增反应；②也可以使用荧光标记的一种三磷酸脱氧核糖核苷酸(dNTP)，如 Cy3 - dCTP/、Cy5 - dCTP/UTP，使它们在 PCR 扩增的过程中掺入到靶基因序列中。目前常使用的荧光物质有荧光素、罗丹明、HEX、TMR、FAM、Cy3、Cy5 等。

与寡核苷酸阵列进行杂交的样品先用不对称 PCR 进行扩增，一般反向引物(reverse primer)与正向引物(forward primer)的比例为 25∶1 ~ 50∶1。同时加入 Cy5 - dCTP 标记扩增产物，使 Cy5 - dCTP∶dCTP = 1∶10。扩增产物用乙醇沉淀或过柱纯化，加入 20 × SSC 和 1% SDS 使终浓度为 5 × SSC 或 7 × SSC，0.1% SDS，混合后煮沸 2 min 变性，室温冷却(防止 SDS 析出)，加样于芯片点阵(1 cm^2 盖玻片约加样 10 μL)，进行杂交。

PCR 扩增后的产物分离的方法也很多。可以用 PAGE 分离、亲合素偶联磁珠(PCR 扩增时，一条引物用生物素标记，另一条用荧光素标记)捕捉扩增产物等。

三、核酸分子杂交和杂交微环境

标记的探针与固定在芯片上的靶分子之间在严格的条件下进行分子杂交，芯片杂交属于固液相杂交，与膜杂交相似，影响异源双链形成的因素包括：①固定在芯片上的靶分子数；②探针浓度；③杂交双方的序列特性；④杂交体系中的盐浓度；⑤杂交时的温度；⑥杂交溶液的 pH 值。一般来说，选择的条件是要使这个反应体系中的绝大多数杂交反应处于最佳条件下，从而使所有的真阳性反应尽可能的体现而假阳性尽可能的减少。

在杂交体系中芯片上的靶分子约 10 倍于探针浓度时，杂交反应为一级动力学方式，此时的杂交速率主要决定于探针浓度，探针浓度每提高一倍，信号将增强一倍，当靶分子的浓度与探针的浓度比偏离这个数值但仍然高于探针浓度时，则杂交反应为假一级的方式。当靶分子浓度低于探针浓度时则杂交反应为二级反应动力学方式，此时固定于芯片上的靶分子数的细小差异将对杂交速率和信号强度产生较大的影响。所以选择一级或假一级反应的杂交条件可以减小和消除系统误差，这对于获得正确而可信的实验结果是十分重要的。现在的芯片制作技术已可以将大量的靶分子刻印在芯片上，并且研究者在芯片技术的应用中均采用过剩的靶分子来进行杂交和分析。

一价阳离子如 Na^+ 的存在可以提高异源杂交双链生成的速度，其原理是 Na^+ 可以屏蔽带负电荷的磷酸根骨架，而后者对于靶分子 - 探针杂交时的碱基相互作用有干扰作用，通常在 DNA 芯片杂交时应用的 Na^+ 浓度为 1 mol/L，倘若杂交温度显著低于异源杂交双链的熔点时，温度对杂交速率有正影响，一般寡聚核苷酸芯片的杂交温度范围为 25 ~ 42 ℃，而 cDNA 芯片

的杂交温度范围为 55 ~ 70 ℃,但在用电离甲酰胺溶液为杂交体系时,则杂交温度为 42 ℃,总之,最佳的温度和盐浓度都需通过测试来确定。

DNA 序列对杂交的影响比较难以控制,特别是对寡聚核苷酸芯片的影响更为显著,它们必须具有相似的解链温度,这样才能保证杂交条件的一致性,这就与寡核苷酸组成和长度有关;即便长度相同,由于 G + C 含量的差异也具有不同的解链温度,导致杂交动力学的差别。与一般的 DNA 杂交一样,GC 含量较高者所形成的异源双链较稳定,由于在一个芯片杂交反应体系中同时有大量的异源双链形成,因此需要找出一个最佳的杂交条件使尽可能多的靶分子与探针之间形成稳定的异源双链杂交物,从而提高杂交的信噪比。这一条件可以通过添加阳性或阴性对照序列而实验出来。对于寡核苷酸芯片,往往添加一些能平衡 AT 及 GC 杂交结合能的试剂,如四甲基氯化胺(TMAC)来提高杂交的信噪比。

杂交后的芯片需经严格的洗涤过程而除去未杂交的一切留物。这也是杂交流程中最重要的一步。图 15 - 13 列出的是常用的 cDNA 杂交流程。

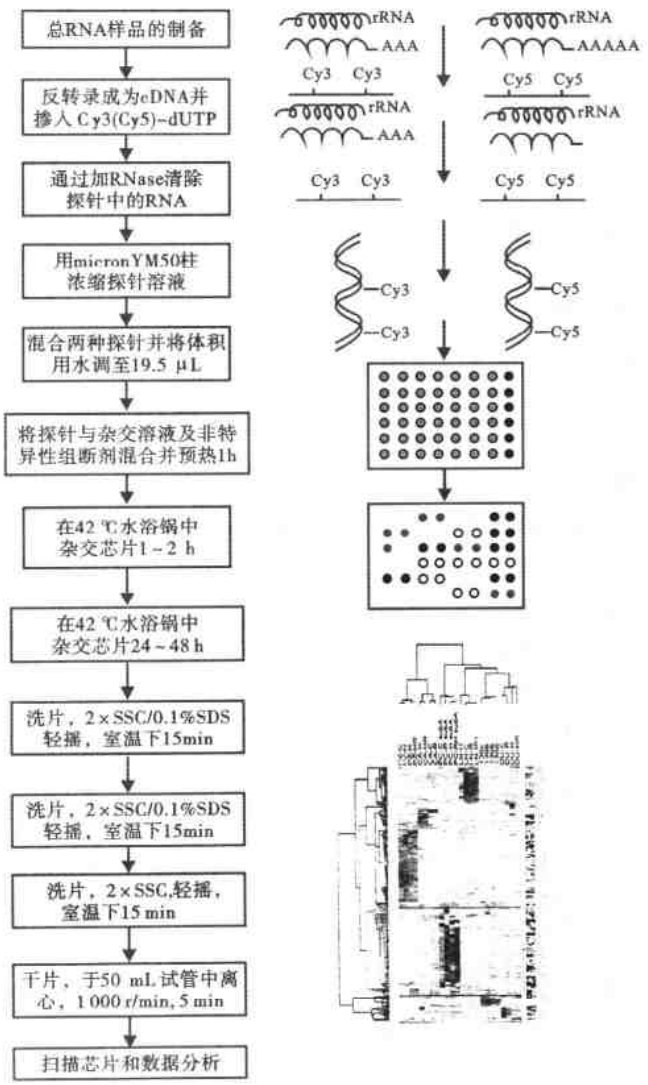


图 15 - 13　cDNA 芯片杂交一般流程图

(彩图见封三)

当然，芯片杂交流程的很多细节会因实验室的不同而有所不同，如杂交所需的总 RNA 通常用 100 μg，但作者在实验中用 25～50 μg 在同样条件下工作得更好，杂交时间 16 h 与 40 h 也没有明显的差异等等。所以，每一个实验者都应根据具体的情况而作出调整以便获得最佳实验效果。

四、DNA 微阵列操作应用实例

前面介绍主要是荧光标记探针及硅芯片或玻璃芯片载体。实际上目前国内广泛使用的还有一类用同位素标记探针的膜载体芯片，技术也较为成熟，重复性也理想，在此也作一介绍。

（一）同位素标记 cDNA 基因芯片的应用实例

以下介绍的实验是应用 cDNA 芯片技术筛选腹主动脉结扎后 4 周大鼠的压力负荷型心肌肥厚相关基因的操作过程。将 18 000 个来自文库的 cDNA 点在带正电荷的尼龙膜上，以质粒 DNA 作为阴性对照验证杂交的特异性。以 8 个看家基因作为阳性对照来标化 mRNA 丰度。样本 mRNA 来自假手术及腹主动脉结扎后 4 周大鼠的心肌组织。在逆转录过程中使用 α－^{33}P－dATP 进行探针标记。然后将探针与微阵列尼龙膜杂交，高严谨度洗涤后经放射自显影获得杂交信号，通过基因信号的相对强度值比较两组间表达水平的差异。结果显示两组之间有 235 个基因片段的信号差别在 3 倍以上。

1. 试剂与仪器。分离 mRNA 的 Kit（QIAGEN 公司）；逆转录 Kit（Promega 公司）；*Taq* 酶及鱼精 DNA（美国 Sigma 公司）；自动点膜仪（BioRobotics 公司）；FLA－3 000 信号记录仪（FIJIFILM 公司），其余试剂均由上海中科院细胞所马普第二实验室提供。

2. cDNA array 的制备（点样法）。首先按常规方法用 PCR 扩增 cDNA、EST 文库，然后通过特殊的针头和微喷头，分别把不同的扩增产物逐点分配在尼龙膜或其他固相基片表面的不同位点上，并通过物理和化学的结合使之固定于相应位点上。点样法的优越性在于可以充分利用原有的 cDNA 文库且长度可以任意选择，固定方法也比较成熟，灵活性大，适于根据需要自行制备点阵规模适中的基因芯片。

（1）挑克隆，摇菌。取－80 ℃保存的单克隆菌种板，将 5 μL 克隆转移至 96 孔深孔板上（每孔含 1 mL LB 培养液），37 ℃摇床（摇速为 250 r/min）中培养过夜。

（2）质粒的提取（操作均在 96 孔板中进行）。

1）将每孔装有 1 mL 菌液的 96 孔深孔板 4 ℃离心（2 000 r/min，10 min）。

2）倒去上清液，并将 96 孔板倒扣在吸水纸上吸干水分。

3）用排枪向每孔中加入 100 μL 溶液Ⅰ，并将沉淀和溶液Ⅰ充分打匀，成悬浊液状。

Solution Ⅰ：	Glucose	50 mmol/L
	Tris.HCl（pH 8.0）	25 mmol/L
	EDTA（pH 8.0）	10 mmol/L

4）用排枪向每孔加入 100 μL 溶液Ⅱ，室温放置至溶液澄清。

Solution Ⅱ：	NaOH	0.2 mol/L
	SDS	1%

5）用排枪向每孔加入 100 μL 溶液Ⅲ，并混匀。

Solution Ⅲ：	5 M KAc	60 μL
	冰乙酸	1.5 μL

H_2O	28.5 μL

6)将各孔中的溶液转移 250 μL 到另一块 96 孔菌种板的各相应孔中，将此板放入离心机，4 ℃离心 (4 000 r/min，30 min)。

7)吸取上清 150 μL 至新的每孔已加好 130 μL 异丙醇的 96 孔板中，稍吹打，放入 -20 ℃冰箱沉淀 15 min。

8)将此 96 孔板放入离心机，4 ℃离心(4 000 r/min，30 min)。

9)弃去上清液，并将 96 孔板倒扣于吸水纸上尽量吸干液体，室温风干至无任何肉眼可见液珠。

10)向每孔加入 30 μL 去离子水溶解。然后各点样 4 μL 进行琼脂糖凝胶电泳，其余样品保存于 -80 ℃备用。

注意：做 PCR 前，每孔要先加入 100 μL 去离子水稀释。

(3)PCR 扩增。

1)每块 96 孔板的 PCR 体系：

H_2O	8.76 mL
10 × Buffer	1 mL
dNTP(25 μM)	80 μL
T3 primer(40 μM)	50 μL
T7 primer(40 μM)	50 μL
Taq 酶	60 μL
总体积	10 mL

按以上顺序将试剂加入到一洁净管中，混匀，然后按每孔 100 μL 的量分装在 96 孔反应板的各个孔中。

2)每孔加入模板 DNA 1 μL，混匀。96 孔板盖上膜后用滚筒压紧，每孔混匀，加水配平后离心。

3)将 96 孔板放入 PCR 仪中，设置好各个温度，进行 PCR 扩增。

PCR 扩增条件：99 ℃ 变性 2 min；94 ℃变性 40 s，57 ℃退火 40 s，72 ℃延伸 2 min 40 s，共 37 个循环；72 ℃ 再延伸 5 min。

PCR 产物离心片刻。

4)每块 96 孔板都需要取 5 μL/孔的 PCR 产物进行琼脂糖凝胶电泳鉴定，以检测是否有 95%以上的克隆被有效地扩增出来。若一块 96 孔板有 5 个孔以上出现扩增失败或含杂带，则该板被认为是不合格的，需重新扩增。

5)然后用排枪向每孔加入 80 μL 异丙醇，混匀。在 -20 ℃冰箱中放置 60 min，然后在离心机中配平，4 ℃离心 (4 000 r/min，30 min)。

6)弃去上清，将沉淀晾干。每孔加 8 μL 变性液，用排枪转移至 384 孔板保存。用作点膜的样品。

(4)点膜。通过自动点膜仪上预先设定的程序，点样针在计算机控制下自动地完成从 96 微孔板中取样、点样、清洗以及再取样等重复操作，直至完成全部微阵列的制作。尼龙膜上可点成千上万个来自人类各个组织的基因，以质粒 DNA 作为阴性对照来证实杂交的特异性，以若干个看家基因作为阳性对照来标化 mRNA 的丰度，每点的 DNA 量约数个纳克。

3. cDNA Microarray 检测的实验流程。

纯化总 RNA

↓

用 *DNase*Ⅰ消化

↓

检测总 RNA 的产量和纯度

↓

富集 polyA$^+$ mRNA，合成探针

↓

探针纯化并检测标记掺入率

↓

与空白膜杂交检测每一探针的非特异背景

↓

杂交反应

↓

结果分析

(1)探针的制作及标记

第一步：分离 mRNA。

提取感兴趣的组织或细胞的总 RNA，以细胞为例，提取过程如下：

1)根据细胞的数量加入适量 Trizol 变性液，每 10^7 细胞加入 3 mL。

2)剧烈振荡以充分悬浮细胞，置冰浴中 5～10 min。

3)充分混匀，4 ℃离心，15 000 g，5 min，转移上清到另一离心管，加入 6 mL 饱和酚，振荡 1 min，放冰浴中 5 min，再加入 1.8 mL 氯仿，振荡 1 min，放冰浴中 5 min。

4)4 ℃离心，12 000 g，10 min。

5)转移上清到另一离心管，重复第 3 至 5 步。

6)转移上清到另一离心管，缓慢加入异丙醇，充分混匀，放冰浴中 5 min，中间摇动数次。

7)4 ℃离心，15 000 g，15 min，弃上清，加入 1 mL －20 ℃预冷的 75%的乙醇。

8)4 ℃离心，15 000 g，15 min，弃上清，干燥沉淀，将沉淀悬浮于无 RNase 的水中，使最终浓度约为 1～2 μg/μL。

9)留 2 μg RNA，用于与经 DNase 处理后的 RNA 进行比较。

10)测量 A_{260}/A_{280}的光吸收比值，纯的 RNA，该比值应为 1.9 到 2.1，在 1%的琼脂糖变性凝胶电泳图上，28S/18S RNA 条带的亮度应该为 1.5∶1～2.5∶1。

(2)探针的制作及标记第二步：DNaseⅠ处理总 RNA。

下列反应均需在无 RNase 的环境中进行。

1)反应体系如下：

总 RNA(1 mg/mL)	500 μL
10× DNaseⅠ缓冲液	100 μL
DNase Ⅰ(1 U/μL)	50 μL
无 RNase 水	350 μL
总体积	1 mL

2)混匀,37 ℃孵育 30 min。

3)加入 10×终止混合液 100 μL,混匀,分成两管,每管 550 μL。

4)每管加入 500 μL 饱和酚及 300 μL 氯仿,振荡 1 min。

5)4 ℃离心,15 000 g, 10 min。

6)转移上清到另一离心管,加入 550 mL 氯仿,混匀,15 000 g,4 ℃离心 10 min。

7)转移上清到另一离心管,加入 1/10 体积的 NaAc 和 2.5 体积的无水乙醇,充分混匀,放冰浴中 10 min。

8)4 ℃离心,15 000 g, 15 min,弃上清。80%乙醇洗沉淀一次,将沉淀干燥后溶于无 RNase 水中,检测 RNA 的纯度和完整性。

(3)探针的制作与标记第三步:纯化组织或细胞的 mRNA。

按分离 mRNA Kit 说明书的操作步骤进行。

1)提取的总 mRNA 加入 1 200 μL 稀释液,充分混匀后,4 ℃离心,15 000 r/min,3 min。吸上清液到另一无 RNase 的 5 mL Eppendorf 管中。

2)入试剂盒提供的 Oligotex 悬浮颗粒 70 μL 至此 Eppendorf 管中,充分混匀,室温静置 10 min。

3)4 ℃离心, 15 000 r/min, 5 min,去上清。

4)加入 350 μL 洗涤缓冲液,在 Vortex 上振荡使沉淀悬浮。用微量进样器吸全部溶液到旋转柱(已放在 1.5 mL 管内)中。

5)4 ℃离心, 15 000 r/min, 30 s。

6)弃去 1.5 mL 管中液体,重新套在旋转柱外。向此旋转柱中加入 350 μL 洗涤缓冲液,4 ℃离心, 15 000 r/min, 30 s。

7)重复步骤 6)一次。

8)将旋转柱放在新的无 RNase 的 1.5 mL Eppendorf 管中。吸 50 μL 已预热(70 ℃)的洗脱液到旋转柱中,吹打数次,使树脂悬浮 ,然后 4 ℃离心, 15 000 r/min, 30 s,保留 1.5 mL 管中的液体。再吸 50 μL 已预热(70 ℃)的洗脱液到旋转柱中,吹打使树脂悬浮。4 ℃离心,15 000 r/min, 30 s,将两次洗脱液合并。

9)将洗脱液转移到另一无 RNase 的 1.5 mL Eppendorf 管中,加 80 μL 异丙醇,混匀。于 -80 ℃放置 10 min。在冷冻离心机中 4 ℃离心 12 000 r/min, 10 min。

10)小心倒去上清,向 1.5 mL 管中加入 100 μL 75%乙醇,颠倒混匀。4 ℃离心,12 000 r/min, 5 min。尽量吸去上清,沉淀于室温晾干。加入 10 μL DEPC 水溶解。取 2 μL 用于定量。

(4)探针的制作与标记第四步:逆转录(RT)和^{33}P 标记。

1)在一无 RNase 的 Eppendorf 管中加入 8 μL mRNA,然后加入 3 μL oligo (dT)引物(0.5 μg/μL)和 2 μL 随机引物(0.5 μg/μL),于 68 ℃水浴中加热 5 min,再冰上骤冷。

2)按以下顺序加入各试剂到上述 Eppendorf 管中,再加入 DEPC 水使终体积为 50 μL,混匀。

5 × RT Buffer	10 μL
RNase Inhibitor	1 μL(40 U)
dNTP(33 mM,无 dATP)	1 μL
AMV 逆转录酶	60 U
α - ^{33}P - dATP	10 μL
(浓度为 10 mCi/mL, > 2 500 mCi/mmol)	

将此 Eppendorf 管放入 32 ℃水浴中温育 30 min 后，42 ℃再温育 90 min；追加 dATP(33 mM) 1 μL 和 AMV 逆转录酶约 20 U，于 45 ℃水浴中温育 30 min。将此 Eppendorf 管煮沸 5 min 变性，立即置于冰浴中，作为杂交的探针。

(5)杂交。配制实验所需的预杂交液 50 mL 和杂交液 50 mL。

预杂交液：6 × SSC

0.5% SDS

5 × Denhardt 试剂

100 μg/mL 已高温变性的鱼精 DNA

杂交液：6 × SSC

0.5% SDS

100 μg/mL 已高温变性的鱼精 DNA

1) 点制好的 7.5 cm × 11.5 cm 尼龙膜一端接触一盘 6 × SSC 溶液的液面，使尼龙膜由下到上完全湿润，然后将尼龙膜浸入溶液中泡 2 min。

2) 镊子将湿润的尼龙膜放入杂交管(周长 10 cm)，加入预杂交液 20 mL，然后将杂交管放入杂交炉中，于 68 ℃温育 3 h。

3) 取出杂交管，倒去预杂交液，但保持膜的湿润。迅速加入 2 mL 杂交液，随即分别加入已高温变性过的两组探针。将杂交管置于杂交炉中，于 68 ℃杂交约 20 h。

4) 取出杂交管，将杂交液小心倒入一 EP 管中。用镊子夹住尼龙膜一角，迅速将其置于一个盛有 250 mL 2 × SSC 和 0.5%SDS 的白瓷盘内，于室温浸泡 5 min，间或温和地摇动。

5) 将尼龙膜转移到一个盛有 250 mL 2 × SSC 和 0.1% SDS 的白瓷盘内，于室温浸泡 15 min，间或温和地摇动。

6) 将尼龙膜转移到一个盛有 250 mL 新配的 0.1 × SSC 和 0.5% SDS 的白瓷盘内，白瓷盘置于 HZS - H 型水浴振荡器中，在水温 37 ℃时振荡 1 h。

7) 将尼龙膜转移到另一盛有 250 mL 新配的 0.1 × SSC 和 0.5% SDS 的白瓷盘内，白瓷盘置于 HZS - H 型水浴振荡器中，在 65 ℃时振荡 1 h。

8) 将尼龙膜转移到一个盛有 250 mL 蒸馏水的托盘中，室温漂洗片刻。

9) 将膜置于一叠吸水纸上，以吸走大部分水分。将膜放到杂交炉(68 ℃)内烘干。用磷屏压片。

杂交结果(见图 15 - 14)。

(6)信号扫描和分析。实验室使用的是 FUJIFILM 公司的 FLA - 3 000 信号记录仪。

膜的信号分析标准：

1)两张膜上的背景亮度值均设为 0。

2)同一张膜上任意两个对照基因的亮度值差异在 2 倍之内。

3)同一张膜上代表同一克隆的两个点亮度差异在 2 倍之内。

4)一张膜上同一克隆的两点亮度分别是另一张膜上对应点亮度的 3 倍以上，可以确信这个克隆代表的基因在这两种样品中有明显的表达差异。

5)有表达差异的克隆的两个强信号点的亮度值都在 10 以上(10 是肉眼可检测的亮度值)。

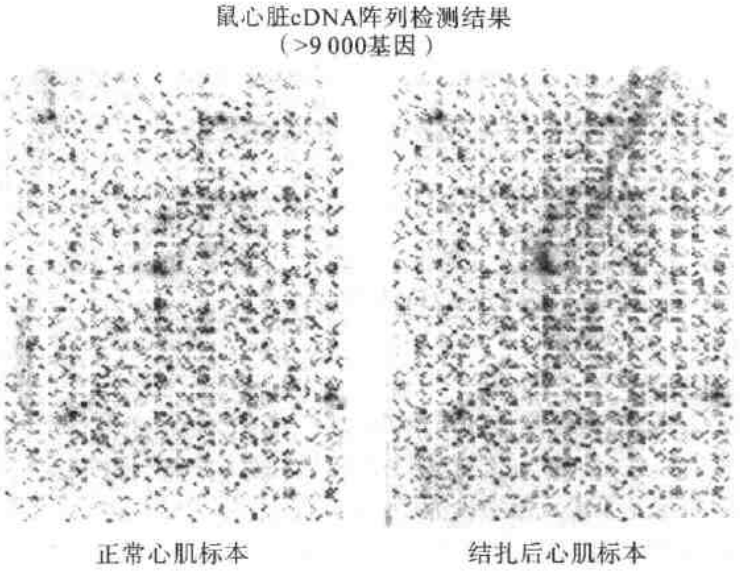


图 15－14　大鼠心肌组织的芯片杂交结果(左为正常心肌,右为结扎后的心肌)

(二)荧光标记 cDNA 芯片杂交实验操作步骤

总体步骤包括:

1. 培养人细胞总 RNA 的制备。

(1)试剂和设备:(文中(a)、(b)、(c)…表示试剂配制方法及注意事项见[附 1]部分)。

①RPMI 培养基(Sigma);②磷酸缓冲液(Sigma);③GTC 溶液;④低速离心机(Backman);⑤超速离心机(Backman);⑥超声清洁离心管,1.3 cm × 15 cm(Backman);⑦CsCl(美国 Biochemical)。

(2)步骤。

1) 取一瓶 50 mL 已培养好的人细胞 (如,Jurkat line J25)。

2) 准备 5 只培养瓶(162 cm^2),每只含 100 mL RPMI 培养基(a)。

3) 以 1:10 稀释分装细胞,并继续培养 48 h。

4) 加入促生长剂(如佛波酯)并再培养 24 h。

5) 将细胞(总量 500 mL)转至 500 mL 离心瓶。

6) 用 JA－10 离心机离心,3 000 r/min × 5 min。

7) 弃上清。

8) 用 10 mL PBS 重悬细胞。

9) 转至 15 mL 锥形离心管。

10) 离心,5 min,1 000 r/min。

11) 弃上清。

12) 将细胞(~ 5×10^8)在液氮中冻干并保存于 －80 ℃。

13) 从 －80 ℃取出细胞(~ 5×10^8)。

14) 加入 12 mL GTC 溶液(b)并分散均匀。

15) 用 12 mL 注射器配以 19 号针头,反复抽吸细胞溶解物 10 次(c)。

16) 准备 6 只超速离心管(d),每只含 3 mL CsCl 溶液(e)。

17) 向每管 CsCl 液面上加入 2 mL GTC 细胞溶解液。

18) 离心沉淀 mRNA。使用 SW50.1 型离心机，在 25 ℃，以 35 000 r/min 速度离心 12 h。

19) 倒置离心管，弃上清(f)。

20) 将离心管倒置于吸水纸上，凉干，60 min。

21) 向每管底部加入 100 μL 1×TE 溶液(g)。

22) 65 ℃加热 1 min，以帮助溶解 RNA 沉淀。

23) 轻轻抽吸混匀 RNA 沉淀。

24) 混和 6 只管中的 RNA 样品混合(共 600 μL)。

25) 用 0.4 mL 新鲜的 1×TE 溶液(g)溶解每管残留的 RNA。

26) 用 0.5 mL 酚(i)抽提 1.0 mL RNA，共两次，以去除残留蛋白质。

27) 用乙醚抽提两次，以去除微量有机物(h)。

28) 将 RNA 样品分装于两只 Eppendorf 管中，每只 500 μL。

29) 每管加入 50 μL 2.5 mol/L 乙酸钠溶液(DEPC 处理过)和 1.0 mL 乙醇。

30) 室温下离心沉淀 RNA 15 min。

31) 弃上清。

32) 在 Speedvac 中干燥 RNA。

33) 用 0.4 mL 的 1×TE 溶液(g)重悬 RNA 沉淀。

34) 重悬溶液的浓度应为 7 mg/mL，即每 500 mL 细胞的 RNA 总产量为 3 mg。

[附 1]

(a) RPMI 培养基须添加 10%胎牛血清，100 μg/mL 链霉素和 500 U/mL 青霉素。

(b) 将 60 g 硫氰酸胍、0.5 g 十二烷基肌氨酸钠、5.0 mL 柠檬酸钠溶于 100 mL H_2O 中，过滤除菌后加入 0.5 mL β-巯基乙醇。

(c) 剪切染色体 DNA 以增加 RNA 产量。

(d) 超声离心管必需清洁。

(e) 将 95 g CsCl 和 20 mL 0.5 mol/L 的 EDTA(pH=8.0) 溶于 100 mL H_2O 中，过滤除菌。加焦碳酸二乙酯(DEPC)至 0.1%。高压灭菌。

(f) 勿使上清回流至 RNA 沉淀中。

(g) 1×TE 溶液。含 1 mmol/L Tris、0.1 mmol/L EDTA(pH8.0)，用 0.1% DEPC 水处理，以灭活 RNase。

(h) 避免管壁内侧含有微量 RNase。

(i) 酚溶液中含酚/氯仿/异戊醇(25:24:1，v/v)。

2. 人总 polyA$^+$ mRNA 的制备。

(1)试剂和设备：(文中(a)、(b)、(c)…表示试剂配制方法及注意事项见[附 2]部分)

Oligotex mRNA Midi Kit(Qiagen)

(2)方法。

1) 解冻贮存于 -80 ℃的总 RNA(7 mg/mL)。

2) 将 150 μL RNA(1 mg)，150 μL 2×结合缓冲液和 55 μL Qiagen Oligotex - dT 树脂混合。

3) 于 65 ℃加热 3 min，变性 RNA。

4) 室温冷却 10 min，mRNA 将退火结合至树脂上。

5) 离心 2 min，沉淀结合有 mRNA 的树脂。

6) 用 600 μL 洗涤缓冲液混悬树脂。

7) 离心 2 min,沉淀树脂。(a)

8) 用 600 μL 洗涤缓冲液重新悬浮树脂。

9) 转入至 Qiagen 离心柱中。

10) 离心 30 s,除去洗涤缓冲液。

11) 在 80 ℃,用 33 μL 洗脱缓冲液重悬树脂。

12) 离心 30 s 并将含 mRNA 的洗脱缓冲液转入一只新试管内。

13) 在 80 ℃,再各用 33 μL 洗脱缓冲液分两次洗脱残留于离心柱树脂内的 mRNA。(b)

14) 合并 3 次洗脱出的 mRNA(每次 33 μL)。

15) 加入 10 μL 2.5 mol/L 乙酸钠和 220 μL 的 100% 乙醇。

16) 在 25 ℃离心 15 min,沉淀 mRNA。

17) 在 Speedvac 中干燥 mRNA 沉淀。

18) 用 11 μL 的 1 × TE 溶液重悬沉淀(pH = 8.0)并混匀。(c)

19) 取 1.0 μL 用于定量。

20) 稀释至 1.0 μg/μL。

21) 产量应为 20 ~ 25 μg mRNA/1.0 mg 总 RNA。(d)

[附 2]

(a) 第一次洗涤时(如采用"多次洗涤")应离心除去微粒和碎片,以免堵塞离心柱。

(b) 在 80 ℃加热装有离心柱的 Eppendorf 管 30 s,以有利于 mRNA 洗脱。该步失败将减低 mRNA 产量!

(c) 溶液用 0.1% 的 DEPC 处理,以灭活 RNase。

(d) 0.1 μL 样品与 0.5 mL TE 溶液混合,于 260 nm 处测吸收度(1.0 OD 相当于 40 μg/mL mRNA)。

3. 用于微阵列点样的 cDNA 的扩增和纯化。

(1)试剂和设备。(文中(a)、(b)表示试剂配制方法及注意事项见[附 3]部分)。

1) PCR 引物用 5′ - 氨基修饰试剂盒 (Glen Research # 10 - 1906 - 90)修饰 C6 臂

2) 96 孔 PCR 仪(PE - 9600 PCR 仪)

3) 96 孔 PCR 板(MicroAmp 96 孔 PE # 801 - 0560)

4) *Taq* DNA 聚合酶(Statagene # 6000139)

5) PCR 纯化试剂盒(Telechem # PCR - 100)

6) 平底 384 孔板(Nunc # 242765)

7) 微点液(TeleChem # MSS - 1)

(2)方法。

1) 向 96 孔板中每孔加入 1.0 μL 质粒 DNA(10 ng/μL)。

2) 加入 99.0 μL PCR 反应液,使反应液最终含有 10 μL 的 10 × PCR 缓冲液(500 mM KCl、100 mM Tris - HCl pH8.3、15 mM Mg^{2+}、0.1% 明胶)、10 μL dNTP 混合物(每种 10 μL)、1.0 μL 引物 1(100 pmol/μL)(a)、1.0 μL 引物 2(100 pmol/μL)(a)、1.0 μL 质粒 DNA(10 ng/μL)(b)、76 μL H_2O 和 1.0 μL *Taq* 聚合酶(5 U/μL)。

3) PCR 扩增 cDNA 30 个循环(94 ℃ 30 s;55 ℃ 30 s;72 ℃ 60 s)。

4) 用 96 孔 PCR 纯化试剂盒纯化 PCR 产物。

5) 用 100 μL 0.1 × TE 溶液(pH8.0)洗脱 PCR 产物。

6) 在 Speedvac 中干燥 PCR 产物。

7) 重悬 PCR 产物于 7.5 μL 微点液中。

8) 转至 384 孔平底板(Nunc)用于点阵。

[附 3]

(a) 整个过程可使用引物(21 mers)。每个引物的 5′末端的 C6 臂要氨基化,PCR 时就可形成连有氨基的 cDNA。

(b) 质粒 DNA 可通过碱裂解法制备和纯化.96 孔 REAL Prep.试剂 (Qiage # SQ811 和 # 19504)可促进快速制备。

4. 微阵列的制备和处理过程。

(1)试剂和设备。(文中(a)、(b)…表示试剂配制方法及注意事项见[附 4]部分)

①微点加样机械臂(Cartesian);②芯片微点加样器(TeleChem);③硅烷化芯片基片(载体)(CEL Associates # CSS - 100C)。

(2)方法。

1) 取硅烷化芯片基片(活性醛基)。

2) 用微点加样器按厂家说明进行带有氨基的 cDNA 的加样。(a)

3) 可在芯片盒中过夜,使点阵上的样品干燥。

4) 用 0.2%SDS 室温浸泡芯片 2 次,每次 2 min 并剧烈搅动。

5) 用蒸馏水室温浸泡芯片 2 次,每次 2 min 并剧烈搅动。(b)

6) 将芯片转至 95 ~ 100 ℃蒸馏水中 2 min,以使 DNA 变性。

7) 在室温彻底干燥芯片(5 min)。

8) 室温下,将芯片转至硼酸钠溶液中 5 min,以减少游离醛基。(c)

9) 室温下,用 0.2% SDS 溶液洗涤芯片 3 次,每次 1 min。

10) 用蒸馏水在室温下再洗涤 1 次,1 min。

11) 将芯片浸入 95 ~ 100 ℃ 的蒸馏水中,30 min。(d)

12) 可凉干芯片并贮存于 25 ℃黑暗处(稳定期 > 1 年)。

[附 4]

(a) 干燥能增加交联效率,可多进行几天。

(b) 此步除去盐和未结合的 DNA。

(c) 将 1.0 g $NaBH_4$ 溶于 300 mL PBS 中。加入 100 mL 100% 乙醇以减少气泡。用前临时配制!

(d) 加热芯片有助于干燥过程。

5. 体外转录法制备对照 mRNA。

(1)设备和试剂。(文中(a)、(b)、(c)…表示试剂配制方法及注意事项见[附 5]部分)

RNA 转录试剂盒(Stragene)

(2)方法。

1) 取异种来源的 cDNA 克隆插入 pBluescript 载体。(a)

2) 用 25 μL 反应体系并含有 poly(A)$_n$ 下游区域单一位点的限制性内切酶消化 10 μg 质粒 DNA。(b)

3) 抽提线性质粒,包括酚 1 次、酚/氯仿/异戊醇(25:24:1 v/v)1 次和乙醚 3 次。(c)

4) 将抽提物置于 65 ℃保温 5 min,并轻缓旋涡振荡,以除去残留的乙醚,加入适量蒸馏水或 1 × TE 溶液。

5) 每 25 μL 线性质粒中,加入 0.1 倍体积的 2.5 M 乙酸钠和 2 倍体积的乙醇,旋涡振荡混合。

6) 室温下离心 5 min,沉淀线性质粒。

7) 弃上清并在 Speedvac 中干燥 DNA。

8) 用 10 μL pH8.0 的 1 × TE 溶液(10 mmol/LTris – HCl 和 1 mmol/L EDTA)重悬线性质粒 DNA,终浓度为 1.0 μg/μL。

9) 在一微量离心管中,加入 20 μL 的 5 × 转录缓冲液、4 μL 线性质粒(1.0 μg/μL)、16 μL 的 4 种 dNTP 混合物(每种 2.5 mmol/L),4 μL DTT(0.75 mol/L),3 μL RNase 抑制剂(20 U/μL)和 52.5 μL H_2O,轻轻混匀。

10) 加入 0.8 μL T3 或 T7 RNA 聚合酶(50 U/μL)(d),轻轻混匀。

11) 37 ℃孵育 30 min。

12) 加入 10 μL 的 2.5 mol/L 乙酸钠。

13) 抽提 110 μL 反应产物,包括酚/氯仿/异戊醇(25:24:1 v/v)抽提 1 次和用乙醚抽提 3 次。

14) 在 65 ℃保温抽提物 5 min,并轻缓旋涡振荡以除去残留的乙醚。

15) 加入 220 μL 乙醇并离心 5 min,以沉淀转录产物。(e)

16) 弃上清并在 Speedvac 中干燥 DNA。

17) 用 10 μL pH8.0 的 1 × TE 溶液(10 mmol/L Tris – HCl 和 1 mmol/L EDTA)重悬转录产物,终浓度为 4.0 μg/μL。(f)

18) 用体外转录的 mRNA 作为基因表达监测对照(见步骤 7,9)。

[附 5]

(a) 克隆应尽可能相异。植物克隆(如 Arabidopsis)适用于人和酵母,反之也然。克隆应带有 20 ~ 40 个核苷酸的 polyA 尾,以便转录反应时与引物 Oligo – dT 互补结合。

(b) 用于体外转录的线性质粒应含有 polyA$^+$ 位点及紧靠其下游的酶切位点(如 *BamH* I),但应尽可能少含载体序列。

(c) 去除微量核酸酶可防止体外转录物降解。

(d) T3 或 T7 的选择,将由聚合酶的性质决定,聚合酶从克隆 cDNA 模板上转录出信息 mRNA 链。信息链作为 top stand,其 3′末端紧靠限制性酶切位点上游所含的 polyA 尾。

(e) 体外转录的效率很高,通过离心可见大量的 RNA 沉淀。

(f) 转录的完整性可通过甲醛凝胶电泳或 mRNA 通过 Oligo – $(dT)_n$ 引物和逆转录生成的 cDNA 碱性琼脂糖电泳来评价。

6. 线性 PCR 法制备对照荧光 cDNA(不对称 PCR)。

(1)设备和试剂。(文中(a)、(b)表示试剂配制方法及注意事项见[附 6]部分)

①PE – 9600 型 PCR 仪(或相当设备);②QIA 快速 PCR 纯化试剂盒(Qiagen # 28106)。

(2)方法。

1) 将异种来源的 cDNA 克隆插入质粒载体。

2) 用 cDNA 特异引物通过 PCR 扩增含异源 cDNA 的重组质粒。

3) 用 QIAquick 柱纯化扩增产物。

4) 在 Speedvac 仪中干燥样品。

5) 用 10 μL pH8.0 的 1×TE 溶液(10 mmol/L Tris－HCl 和 1 mmol/LM EDTA)重悬纯化的 cDNA 重组体。

6) 在一微量离心管中,加入 71 μL H_2O、10 μL 10×PCR 缓冲液(500 mmol/L KCl,100 mmol/L pH8.3 的 Tris－HCl、15 mmol/L $MgCl_2$、0.1%明胶)、10 μL dNTPs(每种 2 mmol/L)、5 μL Cy5－dCTP (1 nmol/L)(a)、2.0 μL 20 mers 寡核苷酸(100 pmol/μL)、1 μL cDNA 的 PCR 产物(0.5 U/μL),轻轻敲打混匀。

7) 加入 1.0 μL *Taq* DNA 聚合酶(5 U/μL)(d),轻轻混匀。

8) 荧光素标记单链 cDNA 按下列条件扩增:95 ℃变性 2 min、然后在 94 ℃ 30 s、55 ℃ 30 s、72 ℃ 30 s 的循环条件下扩增 30 个循环,72 ℃延伸 2 min。贮存于 4 ℃待纯化。

9) 用 QIAquick 柱纯化荧光线性扩增产物。

10) 在 Speedvac 仪中干燥样品。

11) 用 50 μL pH8.0 的 1×TE(10 mmol/L Tris－HCl 和 1 mmol/L EDTA)重悬沉淀,荧光标记的 cDNA 终浓度应为 40 ng/μL。

12) 每 20 μL 杂交液(步骤 7 中第 18 步)中加入 1.0 μL 40 ng/μL 荧光对照,调节荧光对照的 cDNA 浓度为 2 ng/μL。

[附 6]

(a) 可选其他荧光素,如 F12－dUTP、L5－dCTP 和 Cy5－dCTP。

(b) 这种对照可提供一个杂交和扫描的参照,有如逆转录过程(见步骤 7)。2 ng/μL 的荧光标记物应能产生与大量细胞转录同样强的信号。可用含多种荧光的 cDNA 来产生一荧光系列。

7.人总 mRNA 荧光探针的制备。

(1)试剂和设备。(文中(a)、(b)、(c)…表示试剂配制方法及注意事项见[附 7]部分)

1)StrataScript 逆转录 PCR 试剂盒(Stratagene #200420)。

2)Oligo－dT 21 mers (0.1%DEPC 处理)。

3)100 mmol/LdATP、dCTP、dGTP、dTTP (Pharmacia #27－2050、－2060、－2070、－2080)。

4)1 mmol/L Cy3－dCTP(Amersham #PA53021)。

5)1 mmol/L Cy5－dCTP(Amersham #PA53021)。

6)1 mmol/L fluoresecein－12－dCTP (DuPont #NEL－424)。

7)SuperScript II RNase H－逆转录酶 (Gibco BRL #18064－14)。

(2)方法。

1) 在一微量离心管中,加入 5.0 μL 总 polyA$^+$ mRNA(1.0 μg/μL)(a)、1.0 μL 对照 mRNA 混合物(0.5 ng/μL)(b)、4.0 μL Oligo－dT 21 mers(1.0 μg/μL)和 27.0 μL H_2O(DEPC 处理过)。

2) 65 ℃变性 mRNA 3 min。

3) 加入 10.0 μL 5×第一链缓冲液(250 mmol/L pH8.3 的 Tris－HCl、375 mmol/L KCl、15 mmol/L $MgCl_2$)、5.0 μL 10×DTT(0.1M)、1.5 μL RNase 抑制剂(20 U/μL)、1.0 μL dATP、dGTP、dTTP 混合物(每种 25 mmol/L)、2.0 μL dCTP(1 mmol/L)、2.0 μL Cy3－dCTP(1 mmol/L)(c)并轻敲混合。

4) 加入 1.5 μL SuperScript II RNase H－逆转录酶(200 U/μL)使总反应体积为 50.0 μL。

5) 轻敲混匀。

6) 25 ℃退火 10 min,使 Oligo - dT 互补结合 mRNA。

7) 37 ℃逆转录 polyA⁺m RNA 2 h。

8) 25 ℃加入 5.0 μL 的 2.5 mol/L 乙酸钠和 110 μL 100%乙醇。(d)

9) 在 25 ℃下离心沉淀 cDNA/mRNA 杂合链, 15 min。(e)

10) 弃上清并用 0.5 ml 80%乙醇小心洗涤沉淀。(f)

11) 在 Speedvac 中干燥样品,然后用 10.0 μL 1×TE 溶液(pH8.0)重悬沉淀。(g)

12) 煮沸 3 min 变性 cDNA/mRNA 杂合链,迅速置冰上冷却。

13) 加入 2.5 μL 1N NaOH 并置 37 ℃保温 10 min 以降解 mRNA。

14) 加入 2.5 μL 1 mol/L Tris - HCl(pH6.8)和 2.0 μL 1 mol/L HCl 以中和 cDNA。

15) 加入 1.7 μL 2.5 mol/L 乙酸钠和 37 μL 100%乙醇。

16) 全速离心 15 min 以沉淀 cDNA。

17) 弃上清并用 0.5 mL 的 80%乙醇小心洗涤沉淀。(f)

18) 在 Speedvac 中干燥样品,然后用 13.0 μL H_2O 重悬。(g)

19) 加入 5.0 μL 的 20×SSC 溶液(即 pH7.0 的 3 mol/L 柠檬酸钠)和 2.0 μL 2%SDS 溶液。

20) 65 ℃加热 30 s,以溶解荧光混合物。

21) 高速离心 2min,以除去微量碎片。

22) 将上清转至一新离心管内。

23) 探针的终浓度应为:在 20 μL 5×SSC 和 0.2%SDS 中,每种荧光素浓度为 0.25 μg/μL。

[附 7]

(a) 用 Oligo - dT(Qiagen)纯化来自总 RNA 的总 m RNA。

(b) 假定体外转录的对照 mRNA 平均长度为 1.0 kb,按分子比 1:1 000,1:10 000 和 1:100 000 确定浓度。

(c) 可选其他荧光素标记 mRNA,如 F112 - dCTP 或 Cy5 - dCTP。

(d) 冷却或使用 >2 倍体积的乙醇将产生未标记的沉淀。

(e) 由于沉淀附着在管壁的一侧,应从管的另一侧倒去上清。

(f) 为防止沉淀损失,在倒去 80%乙醇之前应先离心 1 min。

(g) 产物沉淀通常附着在管壁的一侧,注意彻底溶解之。

(h) 沉淀溶解后残留的不溶物将导致荧光背景加深。

8. cDNA 微点阵杂交和洗涤。

(1)试剂和设备。(文中(a)、(b)、(c)…表示试剂配制方法及注意事项见[附 8]部分)

1) 杂交盒(TeleChem)。

2) 阵列洗涤池(TeleChem)。

3) 阵列扫描仪 3 000 (General Scanning)。

(2)方法。

1) 将点阵好的芯片放入杂交盒中。

2) 加入 5.0 μL 5×SSC 和 0.2%SDS 溶液进行杂交。(a)

3) 沿 22 mm×22 mm 的覆盖条边缘加 6.0 μL 荧光探针。

4) 用镊子将覆盖条盖在芯片阵列上,在覆盖条和点阵之间形成一个薄薄的样品层。(b)

5) 封闭带有芯片的杂交盒。

6) 将杂交盒置于 62 ℃的水浴中。(c)

7) 62 ℃杂交 6 h。

8) 杂交后立刻将芯片放到洗涤池中。(d)

9) 用 1 × SSC 和 0.1% SDS 室温洗涤 5 min。(e)

10) 将洗涤池和芯片转至第二个含有 400 mL 0.1 × SSC 和 0.1% SDS 的大烧杯中。

11) 室温洗涤 5 min。

12) 用含有 0.1 × SSC 的第三个烧杯简单地洗涤芯片以除去 SDS。

13) 凉干芯片。

14) 用芯片扫描仪 3 000 的激发荧光扫描芯片。

[附 8]

(a) 防止覆盖膜下的样品变干。

(b) 覆盖膜应无油脂、灰尘和其他杂质。

(c) 将覆盖膜从左到右铺到芯片上,以防气泡产生。膜下的小气泡在 62 ℃加热几分钟后即可除去。

(d) 62 ℃最适宜 cDNA – cDNA 杂交。寡核苷酸杂交的温度应稍低些。

(e) 洗涤池应置于含有 400 mL 1 × SSC 和 0.1% SDS 的 600 mL 大烧杯中。芯片应迅速从杂交盒转至洗涤池中。如在室温下放置芯片过久,将导致荧光背景加深。

(f) 在洗涤这一步,覆盖膜应很快脱下,若 30 s 内还未拿走,快用镊子将膜取走,否则会影响洗涤效果。

9. 微阵列用于基因表达分析。

(1)试剂和设备。(文中(a)表示试剂配制方法及注意事项见[附 9]部分)

1) 硅烷化芯片基片(CEL Associates # CSS – 100)。

2) 各种 cDNA 片段。

3) 从 mRNA 转录的样品。

4) 微阵列加样器(Cartesian Technologies)。

5) 阵列扫描仪 3 000 (General Scanning)。

(2)方法。

1) 制备含有目的基因的 cDNA 微阵列。(a)

2) 酶标法制备总 mRNA 的荧光样品(见前)。

3) 按严格的漂洗条件进行荧光样品与微阵列的杂交。

4) 漂洗除去未杂交的样品。

5) 用阵列扫描仪 3 000 的激发荧光扫描芯片阵列。

6) 测量每个点的荧光强度。

7) 获取基因表达扫描值,通过适当对照与实验数据进行比较。

[附 9]

(a)阵列片段可为已知或未知的 cDNA、来自给定基因组的限制性片段或所有基因。来自表达序列(如 ESTs)的合成寡核苷酸也可用于点阵。

10. 寡核苷酸点阵用于突变检测。

(1)试剂和设备。(文中(a)、(b)、(c)…试剂配制方法及注意事项见[附 10]部分)

1) 合成的 15 – mers 寡核苷酸片段,氨基修饰 5′端(Stanford PAN facility)。

2) 硅烷化芯片基片(CEL Associates # CSS – 100)。

3) 2×微点液(TeleChem)。

4) 来自 DNA 的荧光标记样品。

5) UniHyb 杂交液(TeleChem)。

6) 微阵列加样仪(Catesian Technologies)。

7) ImaGene 软件(BIO Discovery)。

8) 阵列扫描仪 3 000 (General Scanning)。

(2)方法。

1) 制备 15 – mers(a)寡核苷酸阵列,其中间位置(第 8 个)的碱基用作区别荧光样品的多态性或突变碱基(b)。阵列制备时,1×微点液中含 10 ~ 100 pmol/μL 的寡核苷酸。

2) 处理芯片阵列,以除去未结合的 15 – mers 的寡核苷酸。

3) 通过 PCR 扩增中间带有多态性或突变点碱基的荧光样品(c)。含有 < 1 000 点阵的样品杂交时,使用 0.1 倍的 100 μL PCR 反应体系。样品杂交前用乙醇沉淀或过离心柱纯化。

4) 杂交前,煮沸样品,变性 2 min (d)。

5) 在 42 ℃,将荧光样品与寡核苷酸杂交 4 h(e)。

6) 按下面介绍的方法洗去未杂交的样品:先用 2×SSC 溶液室温漂洗 2 次,每次 5 min ;再用 2×SSC 溶液室温漂洗 1 次,每次 5 min。

7) 凉干杂交芯片。

8) 设定最大的 PMT 和激光值,以保持线性和减低背景(f)。

9) 用 ImaGene 软件计算荧光强度。

[附 10]

(a) 寡核苷酸必须共价结合在固相支持物表面上。此处我们用带有活性醛基的片基与寡核苷酸的碱基上原有氨基共价结合。

(b) 15 – mers 的中间点或第八个碱基位置是通过杂交来检测单个碱基的多态性和突变的关键位置。例如,野生型为"G"而突变型为"T"时,杂交时,这两个 15 – mers 的寡核苷酸将会对其加以区别;除非野生型 15 – mers 的第八个位置含有"C"或突变型寡核苷酸含有"A"。

(c) 数目超过等位基因 60 个碱基的荧光引物也容易杂交到寡核苷酸阵列上。

(d) 双链荧光样品杂交前必须变性。推荐使用线性 PCR 制备的单链荧光样品。

(e) 杂交温度应比 Tm 低 10 ℃。

(f) 42 ℃适用于 15 – mers 的寡核苷酸。还可根据寡核苷酸的长短来调整温度。

(g) 对于大多数基因型应用,将阵列扫描仪的激光和 PMT 各自设定为 70%和 80%是合适的。

第四节　芯片杂交信号的收集及生物信息的分析

一、芯片杂交信号的收集及处理

生物芯片的信号收集与处理的方法取决于信号扫描的方式。一般均限于光信号的扫描和电信号的点扫描两种模式。

在经典的基因芯片扫描仪中,因为采用的是荧光检测原理,所以基因芯片数据一般进行光采集,即通过激光激发掺入检测点中的荧光生色基团,对荧光报告基团所发出的一定波长范围的光进行收集;然后利用光电倍增管或 CCD 将收集所得的光信号转化为电信号并给出相应的

强度；通过软件将电信号还原成为图形或生物信息的相关数据；最后实验人员可以通过分析软件对所得的数据加以分析处理并给出检测报告。

对于微流控芯片如微缩芯片实验室，其检测一般通过阵列毛细管电泳以及阵列传感器原理的方式，因此采集的信号一般是光或电信号。毛细管电泳给出了核酸或蛋白质泳带的位置，可以通过分子量标记直接对位置和光信号强度进行扫描，根据位置回溯寻址并根据强度进行定量分析；阵列传感器大多数通过电信号的有无强弱来对固定在传感器阵列上的样品定量检测分析，也有采用光纤阵列传感器进行生物芯片研究的报道，由于其本质上为光纤传感器，其检测原理也为光信号的检测。

在各种生物芯片的信号分析处理中，DNA 芯片的信号收集处理过程相比较为复杂，发展的相对完善，研究的也相对深入，并有多种仪器及配套软件产品问世。因此我们着重对 DNA 芯片的信号收集与处理加以介绍。

芯片扫描仪是对生物芯片进行信号收集的关键。对于中高密度 DNA 芯片，必须依赖高精度、高分辨率、高灵敏度的芯片扫描仪，才能够较为完整并且精确的收集标记于检测分子上的荧光生色基团激发后释放出的荧光信号，并完成第一步光－电信号转换过程。商业化的基因芯片扫描仪目前主要为激光共聚焦芯片扫描仪和 CCD(charge coupled device)芯片扫描仪两大类。此外还有用荧光显微镜自己组装搭建的 DIY 扫描装置。

一般基因芯片杂交信号种类有：①荧光标记(单标或双标)的靶 DNA(FITC，Cy3，Cy5 等)检测，如 FITC/Cy3，FITC/Cy5，Cy3/Cy5；② 酶标的靶 DNA(灰度图像分析)；③其他种类的靶分子标记(单彩色或多彩色)，如化学发光、放射性核素标记、电化学等方法的不同。

用不同公司的扫描仪及配套软件分析 Cy3 和 Cy5 两种荧光信号的强度和比值并确定其可信限。芯片扫描中的关键问题主要有：①芯片保持洁净，由于芯片扫描仪有极高的灵敏度和分辨率，污染物会引起噪音和背景。故在芯片制作杂交和清洗过程中，应在超净台或超净室中进行。所用溶液和试剂的纯度应符合规定和过滤除菌；②细胞或组织样品测定时会产生较高的背景噪音；③芯片完成杂交和清洗反应后应立即测定，防止靶分子降解和荧光淬灭；④芯片扫描仪应放置平稳，防止外源性震动。

(一)激光共聚焦芯片扫描仪

激光共聚焦芯片扫描仪采用激光作激发光源，使荧光生色基团产生高强度的发射荧光，光电倍增管检测其光信号，具有较高的灵敏度，可检测每平方微米零点几个荧光分子。由于生物芯片荧光斑点明暗的动态范围较宽，因此要求光电倍增管有较宽的光电响应动态范围，此类扫描仪的动态范围一般可达 216，即 16 位。激光经聚焦后产生极小的光斑，因此具有较高的分辨率，可达几个微米(图 15－15)。其重复扫描的精度主要受 X－Y 移动平台的机械精度、重复性和环境条件的影响。从而提高检测灵敏度。生物芯片中，靶分子常用 2 种或 2 种以上的荧光染料标记，故扫描结果常用两种荧光光密度比值来表示，以减少或消除测定时的干扰和实验误差，提高结果的可靠性。为此商业化的扫描仪常采用 2 种以上不同波长的激光器作为激发光源，最多达 4 种，以激发不同荧光染料标记的靶分子。常用的激光器波长范围可从 488nm 至近红外光区。激光共聚焦芯片扫描仪的产品相对较多，应用也最为广泛，主要有 GSI(general scanning，Inc.)公司的 ScanArray 系列，Axon Instruments，Inc.公司的 GenePix pro 系列，Molecular Dynamics 的 Avalanche 以及 Virtek Vision 的 ChipReader 等。

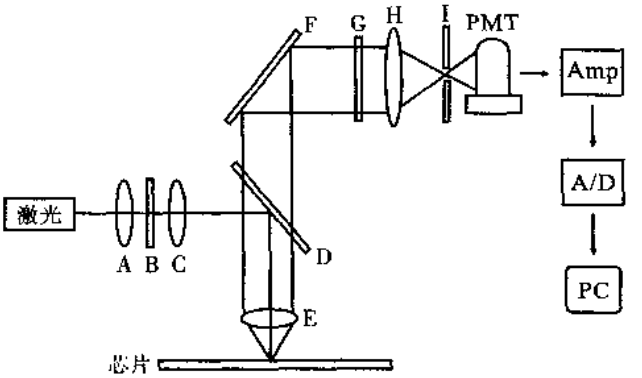


图 15－15　激光共聚焦芯片扫描仪的设计原理示意图

A－透镜组；B－窄带干涉滤光片；C－透镜组；D－二色镜；E－物镜组；F－反光镜；G－窄带干涉滤光片；H－透镜组；I－光阑；PMT－光电倍增管；Amp－放大器；A/D－数模转换器；PC－计算机

扫描结果由计算机采集及处理。扫描得到的图像是黑白的。多种荧光标记的芯片，可分别在不同波长的激光下重复扫描进行；也可在多个激发波长下，同时扫描（即一次扫描可以得多个波长的数据）。后者可克服多次扫描时由于扫描平台机械性能重复性不好造成的误差。

（二）CCD（Charge Coupled Device）芯片扫描仪

相对于激光共聚焦芯片扫描仪，CCD 芯片扫描仪结构较简单，其激发光源多采用氙灯或高压汞灯，单色系统采用激发窄带干涉滤光片。为提高 CCD 芯片扫描仪的检测灵敏度，降低噪音背景，一般采用高强度的氙灯或高压汞灯以提高激发效率，除选择光电转换效率高的 CCD 外，降低 CCD 温度可大大降低 CCD 噪音水平，故采用半导体致冷的冷却 CCD。其温度最低可达－40 ℃，此时 CCD 的灵敏度水平已接近光电倍增管。延长 CCD 成像曝光时间或将多次曝光的图像叠加，也可提高音噪比，提高灵敏度。此类芯片扫描仪的分辨率主要受 CCD 像素总数和被测芯片大小的影响，CCD 总像素应在 100×10^4。但像素的增加会使每个像素感受的光量减少，导致灵敏度的降低。因此 CCD 的选择应与被测芯片的密度相适应，芯片密度高应选择像素较多的 CCD，反之应选择像素较少的 CCD。在 CCD 总像素固定的情况下，芯片越大则分辨率越小。CCD 芯片扫描仪的分辨率一般可达数微米至几十微米。同激光共聚焦芯片扫描仪一样，CCD 测定的动态范围可达 16 位甚至 24 位。CCD 摄像头的镜头前按有电机驱动选择的发射窄带干涉滤光片。生物芯片在经激发窄带干涉滤光片滤光后的单色光激发下产生的荧光，经发射窄带干涉滤光片由摄像镜头捕获成像在 CCD 芯片上。CCD 每次只能读取一个激发波长下的图像，对于多色荧光染料标记的芯片，需要更换激发和发射干涉滤光片，再次读取。由于 CCD 芯片扫描仪是同时读取整个芯片，因此不需要 X－Y 二维移动平台，扫描速度较激光共聚焦芯片扫描仪快，一般仅需 0.5～2 min。商业化的 CCD 芯片扫描仪有 Genomic Solutions 公司的 GeneTac1 000 和 Biotechnology Solutions 公司的 arrayWoRx 等。

其工作原理概括如下：

光源→激发窄带干涉滤光片→芯片→发射窄带干涉滤光片→透镜→CCD 摄像头→CCD 芯片成像→图像转换→数字信号→计算机存储。

二、生物信息的分析和提取

芯片扫描数据的采集、处理、分析和报告是生物芯片检测的重要环节。一张汇集了数十个

到数千个点识别分子的芯片，它们杂交后的光密度值必须及时收集和处理；如果芯片采用多色荧光染料标记，还必须校定和处理各种荧光强度的比例信息，因此，必须有一个专门系统来处理芯片的数据，完整的数据系统应由下列几步骤组成，分述如下。

（一）图像分析和数据提取

芯片扫描仪对生物芯片扫描后得到代表荧光强度的电信号被收集后，通过伪彩处理形成图像文件。扫描得到的图像是数字文件，而不是各样品点的 OD 值、面积、OD 比值等数据，必须经过图像处理提取各样品点的数据，供进一步的生物学统计分析。

图像分析软件功能有：

1. 图像的平滑过滤。去除芯片制作、杂交、清洗和测定过程中的污染、细胞和组织碎片、芯片扫描仪的噪音等所引起的干扰信号。对这些刺峰信号不加以消除将影响实验结果，消除这些刺峰信号的方法较多，如中值过滤法（ImaGene Version1.0，BioDiscovery 公司）、局部消除法、球面消除法（Genepix pro，Axon Instruments 公司）等。

2. 图像背景的确定。图像上各点的光密度值包含了样品的信号和背景的信号，在提取各样品数据前必须将背景扣除。一般将划分区内样品点之外的各像素的光密度平均值作为背景值，也可以由用户在图像中自己定义一块空白区域作为背景，然后计算该区域内各像素的平均光密度值作为背景。对于背景不均匀的芯片图像，每个样品点可以有自己的背景用于计算。一般图象处理先进行平滑过滤处理再进行背景计算。

3. 样品斑点的识别。一般是在图像中选择需识别的区域，输入阵列的行数和列数后，计算机便自动产生一个网格框，套在芯片图像上的每一个样品点上。计算机在框内中心再设置一个圆，使样品点完全落在圆内。圆内信号即为样品信号，圆外信号为背景信号，见图 15－16。对双色荧光标记的芯片，在图像分析前，必须将 2 个图像进行重叠，使两个图像有相同大小的网格和内圆，并定义其中一个为参比图像。

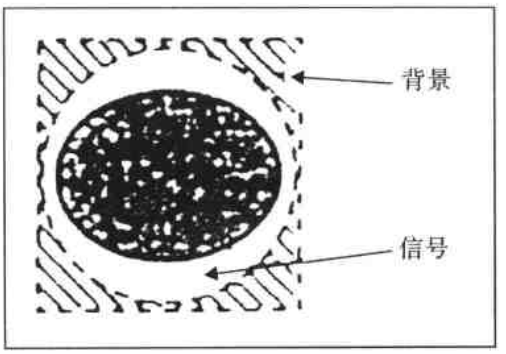


图 15－16　样品信号的确定

4. 数据的提取，存贮与显示。生物芯片中各点阵提取的数据包括：OD 积分值（样品斑内各像素的光密度值之和）、平均光密度值（积分值/样品斑点内像素总数）、样品斑点面积（斑点内像素总数）和样品斑点内光密度最大值和最小值。如双色荧光标记的芯片，还需提取各个样品点中 12 种不同荧光的光密度比值。在由阵列图像转变为数据阵列后，数据提取的任务便完成了，可以储存或输入到特定的数据库中保存。

为直观地观察数据，大多数商业软件都采用伪彩图、散点图和直方图来显示数据。

伪彩图可直观地观察芯片的阵列图像，最强的一般用白色，最弱的黑色，中间用其他颜色如红、橙、黄、绿、蓝等表示见图 15－17。

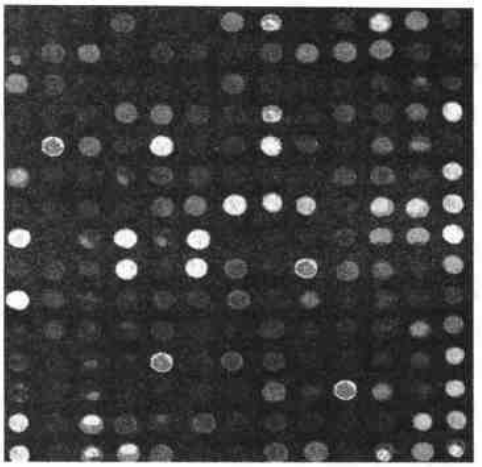


图 15－17　伪彩图

（彩图见封三）

散点图的横坐标表示参比图像的光密度或其对数值；纵坐标表示非参比图像的光密度或其对数值；过原点的 45°线则表示两种荧光标记探针的信号强度相等。用鼠标任击其中一点均可显示出该点对应的 2 种荧光标记的图像，光密度值、光密度值标准差，及表达百分比，见图 15－18。

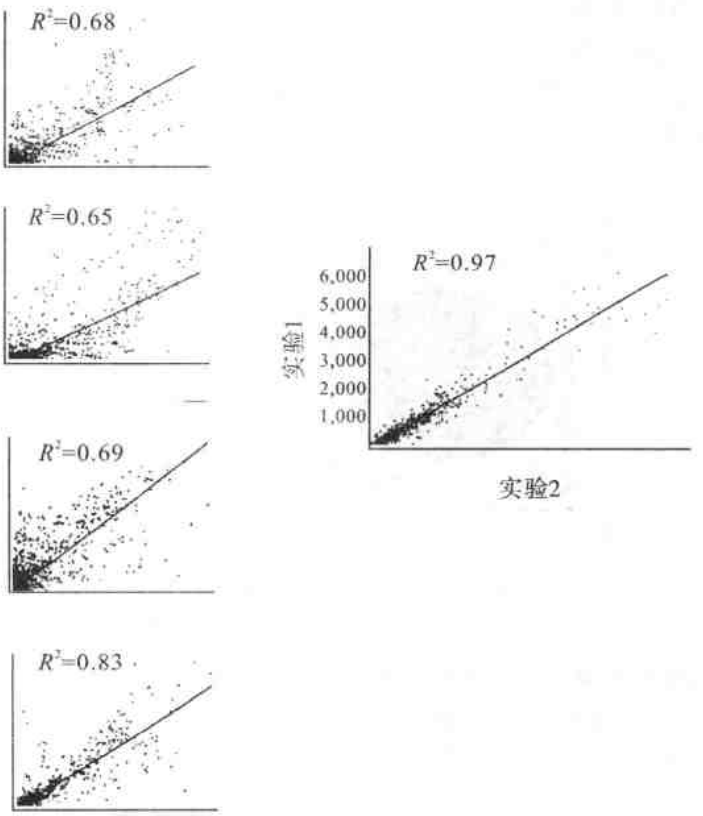


图 15－18　散点图

相同肝脏样品的两个独立 cDNA 微阵列分析散点图。每个点代表 cDNA 簇中的一个基因，相对应的 X 值为一次检测中的表达水平，Y 值表示另一次检测中的表达水平。R^2 为 0.97 表明 cDNA 阵列检测的高度可重复性。

直方图的横坐标表示图像的光密度值;纵坐标表示具有该光密度值的像素的个数。

(二)芯片数据信息的管理

数据库管理的目的是方便地保存、管理和查询数据信息。管理系统应保存下列信息:芯片各点的基因名称、基因的碱基序列或功能的主要描述、GenBank 保存取码、IMAGE 克隆标识、代谢途径标识、内部克隆标识,芯片各点的荧光图像、背景值、光密度值,光密度比值等;还应有芯片的制备、实验、环境条件、样品的类型和制备方法、实验目的和操作者姓名等信息。

(三)芯片数据分析和标准化处理

数据分析是指分析比较芯片各点或芯片间的数据。由于直接比较多个芯片表达的结果是比较困难的,但必须减少或消除各个芯片实验之间的差异。

标准化处理(Array Normalizing)的必要性是由于样本存在差异、荧光标记效率和检出率的不平衡,需对 Cy3 和 Cy5 的原始提取信号进行均衡和修正才能进一步分析实验数据,标准化处理正是基于此种目的。芯片数据标准化处理的方法有多种,常用的有两种:①一般是先确立一组内参照(如一组看家基因),通过内参照来校正生物芯片上所有的样品点阵,包括阴性点,阳性点。杂交时将一定量的与其互补的荧光标记探针混合到杂交液中,这样就可以将对照信号与各样点信号比较,该比值便可消除各实验点间的差异,从而达到归一化的目的。在许多分析步骤中,可以通过标准化来源于每一点阵中的数据来优化比较。有一个标准化方法,它的前提是点阵中的大多数基因都相同地表达。因此在给定点阵中每个特征的算术平均比值就等于1,这个值就是标准化因子。②另一个方法是选择一个位于图像上的特征设置作为控制特征。所有的底物在不同的条件下表达水平都有变化,所以标准化控制特征选择的依据是在点阵中使用的实验条件的一致行为,而不是以前在其他分子生物技术中作为"看家基因"的使用。通过芯片各点数据的分析比较或芯片间的数据比较来实现芯片数据的分析,例如将各点光密度值或比值除以所有点的平均值;或附带一些参数如平均值等以作为该芯片的内部对照。DNA 芯片能同时平行分析大量的信息,包括 SNPs、可表达序列标志(ESTs)和基因克隆、其他有关细胞的基因表达等。

除了上面的标准化方法外,为比较多个芯片的数据,还应严格控制每次的实验条件,如靶 DNA 标记的程序、荧光激发和发射的效率、测定的条件等,使实验在相同的环境和条件下进行。

(四)芯片数据的 Ratio 分析

Ratio 分析是指点阵中每个点 Cy3 的荧光强度与 Cy5 的荧光强度的比值,又称 R/G 值。对于基因芯片而言,一般 0.5 ~ 2.0 范围内的基因不存在显著表达差异,该范围之外则认为基因的表达出现显著改变。由于实验条件的不同,此域值范围会根据可信区间有所调整。处理后的信息再以一种最简单、最直接、最直观的分析方法来表示。通常用散点图(二维或三维)、直方图、饼图、原始图像拼图等。将每个点的所有相关信息如位标、基因名称、克隆号、PCR 结果、信号强度、Ratio 值等自动关联并根据需要筛选数据。每个点的原始图像另存文件,可根据需要任意排序,得到原始图像的拼图,对于结果分析十分有利。

(五)芯片数据的统计分析

目前常用聚类分析,这是一种在事物面孔尚不清楚的情况下研究生物的分类方法。它直接比较样本中各指标之间的性质,将性质相近的归为一类,性质差别较大的归在另一类。可帮助研究人员发现新的基因、新的 DNA 序列、基因的突变位点等。聚类分析根据聚类指标或计

算机方法的不同又可分为许多种方法。其中二向聚类分析(hierarchical cluster analysis)适合高度组织化的基因表达数据,参见彩图 3。

(六)芯片的生物学分析

根据视图分析和统计分析的结果,结合生物学知识对实验结果作出相关判断。例如时间过程分析,是采用较多的一种分析方法,可以得到不同发育时期的基因表达、用药前后基因表达的变化等有用的资料。

(七)GenePix Pro 芯片扫描系统介绍(Axon Inc.)

GenePix Pro 的使用者不需要了解图像处理详细过程就可以获得和分析点阵的图像。尽管GenePix Pro 被用来控制 GenePix 扫描仪,但是也可以使用该软件作为纯粹的分析平台。如果要达到上述目的,则必须用 16 字节的 TIFF(tagged image file format)格式输入点阵图像,才有可能对输入的芯片扫描图像进行分析。

GenePix Pro 的主要参数如下:四波长(可选)同步扫描,扫描精度 5 μm,扫描范围 20 mm × 70 mm。

在讨论图像分析之前,需要一些定义在 GenePix Pro 常出现的一些基本术语。在 GenePix Pro 中,芯片点阵中的每一个单独的点称为特征点,特征点外围均有一环状指示物,能够完全包括特征点即为特征指示物,特征指示物是 GenePix Pro 中分析的基本单位。区域包括了一定范围内的所有特征指示物或所有点,区域指的就是这个范围。在一个点阵中,可有多少个区域,取决于在玻片上点阵的排布方式。

在 GenePix Pro 中要使点阵分析全自动化成为可能,可以使用点阵列表生成器来组建一个点阵列表(GAL)文档,GenePix Pro 通过使用 GAL 文件组建各区域的点阵模式与实际芯片上各区域的点阵相符,并且可以调整各区域的尺寸大小以及位置以适应打印,以及给各个特征点命名。分析过程中关键之处在于每个点的识别过程的得以进行,即区域中每个点所代表含义的保留,因此当实验人员知道点阵中每个点所代表的 IDs 身份和分子生物学数据库的 GI 号码时就可以使用 GAL 文档进行导入,便于以后的数据分析。

许多商业化的点阵仪都自动生成 GAL 文档,GenePix Pro 可以使用装载点阵列表的命令直接向其中装载 GAL 文档。

如果扫描的 DNA 芯片有条形码,GenePix Pro 可以自动地从前面的扫描和保存图像读取条形码符号。拉长的条形码符号在 GenePix Pro 窗中口底部的"B"区报告出来,在结果文档的头部输出。另外也可以用条形码符号作图像的文档名。

GenePix Pro 读取 4 种最普通的条形码,即 39 码、128 码、I2 of 5 码和条形码。这 4 种码的特征如下表 15 - 2:

表 15 - 2 四种条形码特征

条形码类型	数值	α值	特殊值	紧缩值
39 码	0 ~ 9	A - Z,空格	— $ % + /	不限
128 码	所有 ASCⅡ值			
I 2 of 5 码	0 ~ 9	N/A	N/A	不限
条形码	0 ~ 9	N/A	- :. / + $	不限

如果你知道你的条形码的类型，你就可以设置条形码选项使之符合你的需要，这就加速了选择的进程。打开选项/工作，接着单击条形码按钮即可。

1. 图像显示。CenePix 扫描仪采用是双激光同步扫描系统，它可以克服多次扫描带来的二维平台移动的重复性不好所带来的误差。在短时间内产生一个预扫描伪彩图像。这个伪彩图像分辨率较低，也不能够用于图像的分析。在预扫描伪彩图的基础上，实验人员可以选取感兴趣的区域，针对这个区域进行精细扫描。精细扫描得到的图像分辨率较高，并且可以用于进一步进行芯片信号的分析处理。伪彩图像的给出以比值为基础。

在 GenePix Pro 中每一原始波长图像比如绿光 532 nm 或红光 635 nm 图像，都是以 16 字节形式收集的，GenePix Pro 比值图像是标准的 24 字节组成的 RGB(red - green - blue)图像。

当用两种颜色模式来设置错差红 - 绿显示方案和错差比值公式时，来自 532 nm 激光的图像放置在绿色道，来自 635 nm 激光的图像放置在红色道。蓝道设置为 0。每种 RGB 图像的色道都仅仅是由 8 个字节的图像组成的。为了仅仅达到展示图像的目的，每一波长图像必需减少到 8 个字节以便显示 24 字节组成型的 RGB 重叠图像。有两种选择来减少每个原型波长图像到 8 个字节：一是把 16 字节压缩成 8 字节，二是选择 16 字节中的 8 字节。GenePix Pro 通过使用颜色菜单中的命令来提供这两种能力。

为了把 16 字节压缩成 8 字节，软件使用一个平方根转化，图像强度的平方根通过计算机处理显示出来。使 16 字节有效地转化为 8 字节。为了选择 16 字节中的 8 字节，GenePix Pro 可以识别 8 字节并保存它。那些被用来显示比值图像强度范围的选项的效果归纳在下表 15 - 3 中。

表 15 - 3　比值处理方法及图像强度范围

比值处理方法	图像强度范围
平方根转换	0 ~ 65 535
高强度阈值	256 ~ 65 535
中强度阈值	16 ~ 4 095
低强度阈值	0 ~ 255

比值处理方法仅仅影响比值的显示。它对原始波长图像没有什么影响，对显示在结果列表中的比值数据的计算机处理也没有什么影响。下面这个例子就是比值处理的最好证明。假如有一个像素，绿光值是 1 500，红光值是 3 000，如果你的处理比值为红光值比绿光值，那么比值为 2。如果你用的是平方根转换方法，你可以显示从 0 到 65 535 的所有像素。这样值得讨论的是红色像素，这因为红光对 RGB 图像强度的贡献是绿光的 2 倍。

然而当你选用“低强度阈值”时，最大像素就为 255，在这种情况下由于每个像素都被当作是 225，所以红光和绿光对 RGB 图像的贡献就相同了。结果就是一个黄光的比值显示，因为它符合比值为 1(225/225)。

生物信息通过比值图像传递的特征点的颜色以及荧光强度表达出来，比值图像典型地代表了同底物相联系的荧光水平。假设参照底物标记为 Cy3(它是被波长 532 nm 的绿光激发的)，检测底物标记为 Cy5(它是被波长 635 nm 的红光激发的)，两个底物是混合的而且同载玻片上所点样品相结合。在所获得的比值图像中，红点表明检测底物的特征值比参照底物的特

征值要丰富，这表明检测底物的活性比参照底物的要高。绿点表明检测底物比参照底物的活性要低，黄点表明几乎没有差异。

为了准确量化活性的变化，必需从比值图像中提取给定特征点的强度，而且在计算机处理最终比值前必须精练这些数据。另外其他几个参数也需要测定以便评价给定特征数据的质量。

2.背景消减。首先是计算背景强度，为了减少比如载玻片自发荧光的非特异荧光效应或者同参照及检验底物的非特异结合效应，GenePix Pro 提供了几种背景消减方法，使得在一些比值计算之前背景强度已从特征强度消减了。除了计算背景强度外，GenePix Pro 还可报告那些超过背景相关阈值的像素的百分数：如比背景标准偏差大 1，大 2 的像素的百分数。这些与阈相关的值在评价数据的质量时是有用的。在局部背景消减方法中，对每一特征指示物的不同背景是要计算的，而且还要报告出背景像素强度的平均值，中间值和标准差。对于需要背景消减的所有比值计算来说，中间背景值是有用的。哪些像素用来计算局部背景呢？首先是特征指示物中央的环状区，这个区域的直径是特征指示物直径的 3 倍。这个区域的所有像素都用来计算背景，但有下列情况之一则会造成结果存在偏差。①被计算的像素的残留部分位于邻近的特征指示物中；②被计算的像素没有完全位于特征指示物周围的两个像素大小的环内；③像素位于荧光强度较高的特征指示物内。同时 GenePix Pro 也提供了几种球面背景消减方法。在一个球面方法中，单个背景值用于每一波长的整个点阵中。上面局部方法中已解释过，球面背景值是基于局部背景区域。例如，当计算一个所有特征背景中间值的平均值时，平均值来自所有局部中间背景值。

在一定的区域排列和背景阈值情况下，一个特征点如果具有如下情况之一，则显示“找不到”：①特征点所完全包含的像素数少于 6 个；②特征点直径大于 3 倍的特征指示物直径，或引起同邻近特征点重叠；③特征点直径位于选项/分析设置的界限外；④特征点处于一个同邻近特征点重叠的位置。

如果检测过程中，在特征指示物中的点大于原先的特征点，那么可能是由于极端背景的变化，即明背景像素可能包括特征像素。为了避免这些，应尝试增加分析表格的选项对话框中复合像素的强度值，使它比邻近最明亮的背景像素稍微亮些，以避免软件在背景消减和标准化过程中将特征像素作为背景像素处理，有利于后来的量化比较分析。

3.比值分析。GenePix Pro 计算大量的不同比值，每一比值都相当提供一个对原始数据的评价与检验。这些比值可以分成三个不同的类别：即比值来源于总特征值，即总特征值的以像素密度为单位的算术平均值与几何平均值，例如算术平均值的比值，几何平均值的比值等；比值来源于像素强度比值的如衰退值比值，比值的算术平均值等；比值来源于优化因素的如几何平均值的总和分析，比值的对数分析等。所有的比值和导出的数值均消除了背景的影响，减去了背景值的几何平均值。

4.数据标准化。从不同的点阵进行实验比较数据并非易事。在许多步骤中一些能够产生点阵图像的微小变化都可以引起点阵间难以比较的问题。变化可能是由于标记效率不同，不同染料的化学特性不同，针尖、玻片、扫描设置等的不同而引起的。这些变化因素中的一些可以被标准化校正，但还没有一种合适的标准化方法可以校正所有的变化类型。通常实验技术人员会通过选择在特定实验中应当了解的、能够预知错误源的或知道实验特性的标准化方法来进行数据的标准化处理工作。一般的标准化处理依赖于实验技术人员的经验，例如，有经验的实验人员能够通过修正染料标记的差异来检验通道平衡。

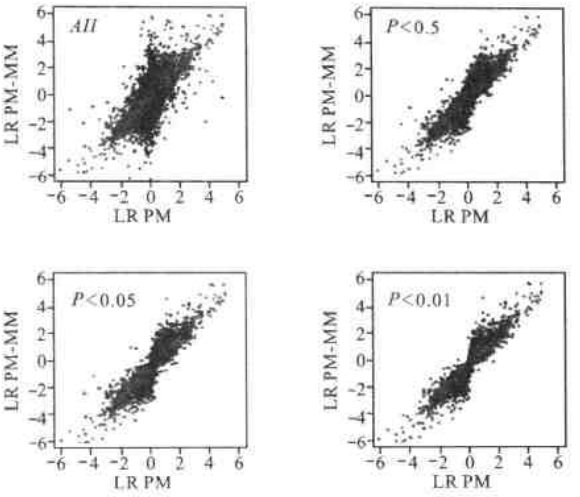
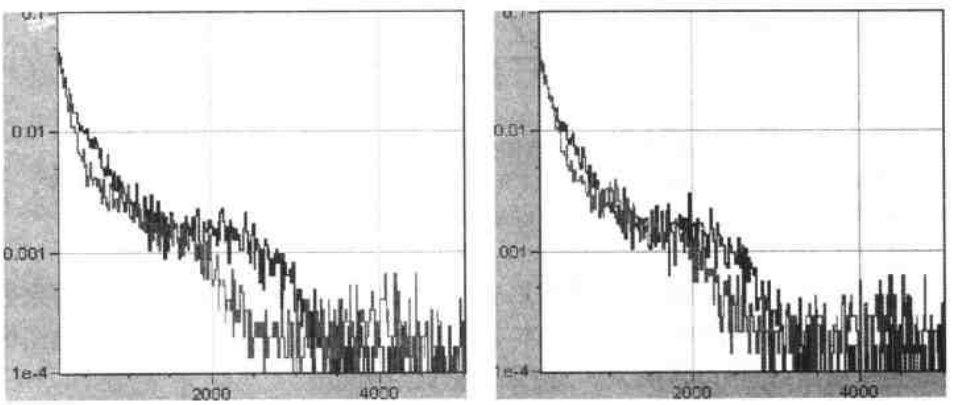


图 15－19

PM(perfect match)与 PM－MM(perfect match mismatch)法比率值之间的比较。图中给出的仅为用 PM－MM 法得出的 P 值小于指标值的那些基因的点，其中最亮的红点为最小的 P 值(约 10－6)，而黑点代表 P 值为 1(左上图)或 0.01(右下图)。

在一个实验步骤中，点阵间变化的主要来源之一就是没有正确设置 PMT 值，因此在一个通道中获得的全部信号同其他道中获得的全部信号就显著地不同，在这种情况下比值就会产生偏差。为了减少这种情况，必需使用初扫描来调整 PMT，以便它们在所有通道中产生相同的反应。(图 15－20)



标准化之前的双色扫描柱状图　　标准化之后的双色扫描柱状图

图 15－20　此图显示近似于真实样本的假设例(以 log 值显示)

标准化前，计算此区域的 Ratio 为 1.12，标准化后为 1.0。标准化前，曲线明显分开，标准化后，曲线大体一致

GenePix Pro 提供了两种数据标准化的方法，一个标准化方法的前提是点阵中的大多数基因都相同地表达。因此在给定点阵中每个特征的算术平均比值就等于 1。这个值就是标准化因子，通过这个标准化因子进行矫正。另一个方法是设置一个位于图像上的特征点作为控制特征。由于所有的底物在不同的条件下表达水平都有变化，所以标准化控制特征选择的依据是在点阵中使用的实验条件行为是一致的，而不是以前的“看家基因”的方法。举个例子说，控

制点可能是每一个点希望得到一个为 1 的比值,因此控制特征的平均值为 1。假设点阵中的变化是相同的,单个标准因子就可以从这些特征中计算出来并应用于整个点阵。

选好了一个标准化方法,必须应用于软件分析中去。GenePix Pro 应用一个球面标准化方法。这个方法中图像和结果都是标准地结合在一起。如果当前结果不是来源于当前图像,那么结果数据只是单独地标准化,你可以通过去掉标准化了的而回到原来的图像。由于比值不是常规地给出,GenePix Pro 在标准化时首先采用每个比值的对数值,比值小 0.1 或大于 10 的排出在计算之外。四颜色的标准化是通过比值与比值间的比值而进行的。例如,如果选择要标准化,那么中间值比值的平均值就设置为 1。在选项/分析/比值公式对话框中定义的中间数据的比值都应单独地标准化。

在 GenePix Pro 中没有系统聚类分析等聚类分析的功能,而只是给出了散点图(scatterplot)分析的功能,散点图只是对分析结果给出一个评价。

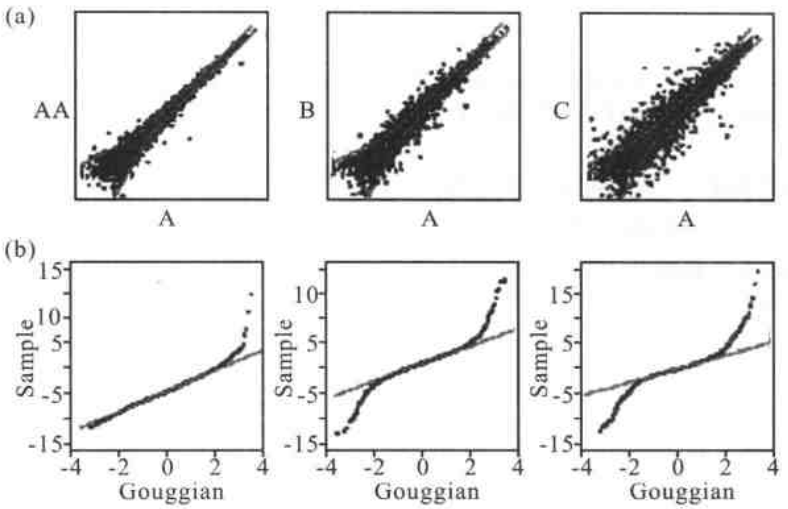


图 15 - 21　散点图分位数 - 分位数(QQ)点

(a)显示微分校准增加水平的典型散点图及 2 个噪声套(红线);有三个不同条件:A、B、C,AA 为 A 的加倍。

(b)LR/σ(η)相关联的 QQ 点。显示取自 12 个 Mu11KsubA 芯片中 4 个的数据

5.分析报告的给出。GenePix Pro 能够以多种格式给出分析结果报告,首先结果数据被保存为 * .gpr 文件,该文件可以被转换成文本文件(* .txt)和 excel 表格文件。扫描图像可以以 * .tiff 格式直接给出,也可以转换为 * .jpg 图形格式给出。

(马先勇　王　震　曹慧青　李　健
文继开　汪世华　刘　强　陆红兵
杨渝珍)

第十六章

微阵列探测中的共聚焦扫描技术

内容提要

激光共聚扫描技术是近20年来发展起来的一种具有特殊检测功能的技术，把该技术应用到微阵列探测上，结合激光、显微镜和计算机，建立一套新型、高精显微扫描系统。该系统利用激光点光源有效地排除焦点以外的荧光干扰，有利于对样本深层结构的探测，实现高分辨率的二维光学断层成像。微阵列探测结合共聚焦技术，具有成像分辨率高、能层析扫描，能观察样品的三维图像等优点，是生物、医学研究领域的一种强有力的仪器，已在生物学领域得到了广泛的应用。随着技术的发展，它们的发展及其应用有着广阔的前景。

2000年6月26日，美、英、德、法、日、中6国科学家同时向全世界公布，人类基因组工作草图绘制完毕，使原本计划于2005年完成的人类基因组计划奇迹般地提前5年完成。这也意味着生命科学研究将从以结构基因测序为主的“结构基因组学”时代进入以破译、识别、解读、开发基因组功能为主的“功能基因组学”时代，也即后基因时代。同时以作图为基础的基因分离将转向以序列为基础的基因分离；从研究疾病的起因向探索发病机理转移；从疾病诊断向疾病易感性研究转移。因此许多国家已开始考虑在后基因组时期，是否能找到有效的硬体技术来对如此庞大的DNA信息加以利用。例如，美国国立卫生研究院、能源部、国防部等机构已拨出了巨额专款，鼓励学术界和商业界对这方面的技术进行开发。面对分析人类基因组10万个基因的复杂表达方式、确定细胞生命中错综复杂的调节表达信号的艰巨任务，传统的一次只能研究少量基因的方法已明显不足，因此就催生了一种新的、革命性的技术——微阵列技术(Microarray technology)。

微阵列技术主要是指由成千上万个DNA样品或寡核苷酸，密集排列于硅片、玻片、聚丙烯或尼龙膜等固相支持物上，再与模板在严格条件下进行杂交，最后通过特定的仪器，如激光共聚焦扫描或CCD(Charge Coupled Device)等设备获取图像信息，通过计算机分析处理获得微阵列信息的技术集合。微阵列又常称为微排列、微点阵或生物芯片(biochip)，目前新兴的DNA芯片(DNA chip)即是微阵列中最主要的一种，通常微阵列与DNA芯片两者的名词常混用。

微阵列(microarray)根据存储生物信息的类型，分为组织微阵列(tissue microarray)，寡核苷酸阵列(oligonucleotide microarray，又称DNA微阵列或DNA芯片)和cDNA微阵列(cDNA microarray，又称cDNA芯片)等。

组织微阵列是大量组织样本高密度排列而成。先在石蜡块上打出间距 0.1 mm、直径 0.6 mm的孔，用针得到直径 0.6 mm 的组织块，再挤压到石蜡孔里面，然后切出 5～8 μm 的切块贴到玻璃片上，经处理后用已知染料进行双色荧光原位杂交。Lukas Bubendorf 等将 371 例前列腺原位癌、复发癌和转移癌样本制成组织微阵列，用荧光原位杂交监测 5 个基因的扩增情况。结果显示组织微阵列技术是大量样本检测肿瘤进展各期分子变化的强大工具。

寡核苷酸阵列用照相平版印刷术(photololthography)和固相合成术结合在载玻片上生成寡核苷酸阵列。由 Fodor 首先在 1991 年报道。该技术是先在玻璃上涂上光敏化学材料，盖上罩，根据所要合成的碱基序列来决定罩的透光位点。核苷酸的 5′末端修饰以光不稳定的保护基团，该基团经光照而去保护，与下一个核苷酸结合。如此重复直至获得所需核苷酸长度。每一层各点的 A、T、C、G 不同，故每合成一层核苷酸需要 4 个不同透光位点的罩和 A、T、C、G 分别冲洗一次。通过 $4 \times n$ 次步骤可合成含有 n 个核苷酸的寡核苷酸阵。其主要的特点是可按需要设计一定序列的寡核苷酸阵链。美国的 Affymetrix 公司已能生产 40 万个寡核苷酸阵/1.6 cm^2的芯片。

cDNA 微阵列最先由 Schena 开发出，它将特定的或文库的 cDNA 经 PCR 技术扩增后，用机械手点到载玻片上。美国 Synteni 公司的 cDNA 微阵列已达到 1 000cDNA/3.6 cm^2。

微阵列自 1995 年在 *Science* 报道后被认为是该年度 *Science* 上发表的最有影响的文章之一。微阵列是新出现的分子生物学技术 、是本世纪重要的科学进展，它能够高效率、大规模地获取相关生物信息，是现代生物技术、微电子技术、机械制造技术、计算机技术的结合。其对科学的深远影响将远胜过 DNA 测序和 PCR 等，使人们更大规模地获取生物信息，使人类“功能基因组计划”早日实现。

生物芯片技术成为人们有效大规模获取生物信息的有效手段，芯片扫描仪是生物芯片能否得到广泛应用的关键仪器，它是利用激光照明生物芯片激发荧光，并用探测器探测荧光强度，以获取生物芯片信息。生物芯片扫描仪主要有使用光电倍增管探测器的激光扫描仪和 CCD 扫描仪。而把共聚焦扫描技术应用微阵列上，不仅能发挥共聚焦扫描光学检测的众多优点，而且开创了微阵列探测的新方法，为人们系统、全面的探索基因功能提供了一种有力工具。

激光共聚焦扫描技术是一项新技术，是经过几十年的不断改进才逐渐发展起来的。1957 年，Marvin Minsky 在其美国专利上首次阐明了共聚焦扫描技术的基本原理。因当时没有合适的光源，Minsky 的仪器产生的图像质量不太好，因此没有引起人们对共聚焦显微的兴趣。十年后，Egger 和 Patran 第一次成功地用共聚焦显微镜产生了一个光学横截面，他们所用的共聚焦显微镜的核心是 Nipkon 盘，此盘位于光源和针孔之后，并且从盘射出的光束以连续的光点在盘旋转时照在物体上。由于 Nipkon 盘在设计中同样受到技术不完善的干扰，也没有获得太大成功。随着计算机、激光和图像处理技术的飞速发展，在 1985 年，Wijnaedts Van Resaint 第一次成功地获得了用荧光探针标记的生物材料的光学截面，充分证明共聚焦荧光显微镜具有摈弃离焦光线，获取生物样品的三维(3D)数据的能力，发表了第一篇有关激光扫描共聚焦显微镜在生物学中应用的文章。在随后的两年内，Oxfordshire 公司，也就是现在的 Bio－Rad 公司生产出第一台商业的激光共聚焦扫描显微镜。随后 Zeiss、Meridian 等多家公司相继开发出各种型号的共聚焦显微镜产品。目前 ，激光扫描共聚焦显微技术已用于细胞形态定位、立体结构重组、动态变化过程等研究 ，并能提供定量荧光测定、定量图像分析等实用研究手段 ，结合其他相关生物技术 ，在形态学、生理学、免疫学、遗传学等分子细胞生物学领域得到广泛应用。

研究表明：共焦扫描系统不仅可以抑制成像中弱杂散光的影响，而且在相同成像条件下，

其分辨力很高。共聚焦生物芯片扫描仪就是集共焦原理、高速扫描技术和计算机图像处理技术于一体的高新技术光电仪器。本章基于此展开讨论，首先介绍微阵列的相关特性及与荧光探测相关的基本知识；第二节讲述微阵列样品制备；第三节介绍共聚扫描原理及共聚焦扫描技术的优点；第四节探讨共聚焦扫描仪的设计和关键特性；最后一节将介绍一种商品化极高的激光共聚焦荧光扫描仪装置(ScanArray® 3 000/5 000)的性能及特点。

第一节　微阵列技术、荧光及荧光探测

所有微阵列上的荧光须要经过扫描装置来分析其上的荧光强度与分布，在这些装置中，激光共聚焦扫描仪具有优越的性能，能获取高质量的图像和数据，显示出优良的性能。在介绍共聚焦微阵探测之前，我们先介绍与扫描设备密切相关的微阵列知识和荧光知识。

一、微阵列探测特征

这里将局限于与荧光探测有关的微阵列知识。微阵列就是许多微小 DNA 或其他生物分子整齐排列在固相表面，这些 DNA 或生物分子与荧光分子结合，测定荧光强度后可推测知结合分子的浓度。受背景荧光限制，典型的荧光动态范围大致在 400:1 ~ 4 000:1 之间。固相载玻片通常是经过化学处理过的玻璃，经常做成 25 mm × 75 mm 大小的载玻片。

在微阵列上包被的分子常常能与多种荧光探针通过化学健相互作用而联结，通常情况下能联结两种荧光分子，例如在检测不同样品的基因表达区别时，可将其中一种样品(如正常组织)标记上发绿光的荧光分子，而将另一种样品如肿瘤组织或发病组织标记上另一种发红光的荧光分子，用两种波长的激光激发微阵列上的样品，通过比较两种荧光在微阵列上某点的比例可以获知某基因在两种组织中的差异。因此，阵列扫描仪通常至少需要两种不同的探测通道。

微阵列上有样品组成的点阵，除了样品点之外，余下的是背景。样品点就是样品，而背景就是点阵周围的区域。如果玻璃载玻片未经表面处理，样品在玻片上结合不牢固，且容易扩散，造成背景高。若经过化学处理后，样品点在微阵列上结合牢固，样品点直径小，不扩散，从而提高样品点的密度。在进行荧光数据分析时，仪器将样品点上的荧光强度与样品点周围的背景强度相比较，如果背景中含有与荧光波段相一致的物质，则背景强度增高，样品的荧光信号将减弱，干扰信号的读取。

微阵列样品点的直径通常在 25 ~ 500 μm 之间，目前所能达到的多数 DNA 微阵列样品点的直径是 100 ± 50 μm，并且朝着更小直径的趋势发展。因为样品点直径的变化对扫描结果的读取产生影响，如果点的直径大，荧光分子扩散的范围也大，则观察到的荧光强度相对较弱。

若微阵列载玻片上有灰尘或其他杂质，如衣服纤维、皮屑、指纹等都将影响扫描结果。各种组织物质能发荧光，微阵列灵敏度应有足够高的灵敏度，能清楚的探测到空气中一般污染颗粒，如衣服纤维、皮肤或指甲油。

“印”在微阵列上的样品点干燥后会形成很薄的层，通常不超过 10 μm。如此薄的条件使得共聚焦扫描成为可能。精心地调整扫描仪的焦距，可避免检测出不需要的背景杂质，包括微阵列上的灰尘、玻璃中自带的荧光杂质产生的干扰信号均可通过激光共聚焦的方式将其减少到最低程度。因为杂交因素影响，在微阵列上样品点与样品点之间的荧光强度差别很大。对微阵列扫描仪的要求要能够有较宽范围的灵敏度，通常来说，灵敏度调整范围为 10 000:1。

二、荧光基础理论

在生物探测中荧光知识范围很宽，其他地方会有详细的描述。这里简要讲述一些与微阵列扫描相关的一些重要荧光概念、知识，并讨论了微阵列探针选取及应注意事项。

1. 特定光谱。荧光团有特定的吸收光谱和发射光谱，荧光团吸收一定波长的光子使基态电子跃迁到更高能态，这将意味着它具有发射荧光的潜能。当受激发的电子回到基态时，能量以发射特定波长的光子而释放。

2. 激发光谱。激发光谱的绘制过程是，在单一波长下，变换激发波长在不同波长下测定荧光强度发射。

3. 发射光谱。发射荧光曲线的绘制过程则是在最长激发激发波长下开始增加波长，测定不同波长下的荧光发射强度。

4. Stokes Shift。染料或荧光团被激发光照射时就会发射荧光，用于微阵列的荧光发射波长通常比激发波长要长，荧光团发射光谱相对激发光谱总有更长波长的偏移，如图 16－1。这种在最大发射峰和最大吸收峰处波长的差别称之为 Stokes shift，也称红位移。

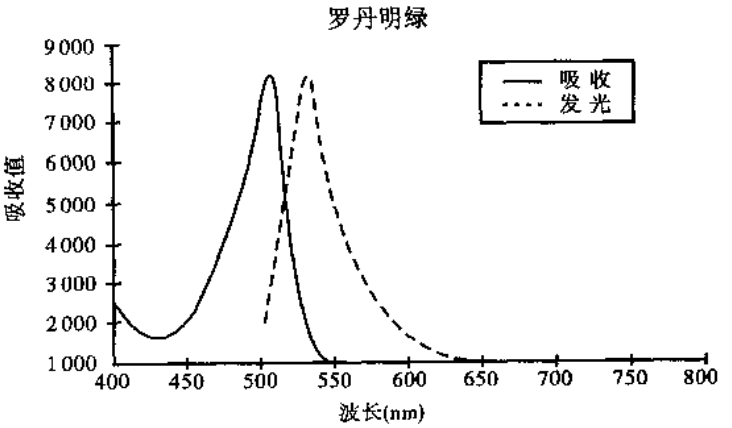


图 16－1　罗丹明绿(Rhodamine Green)荧光光谱

5. 激发态寿命。在短暂的激发态寿命期间，激发态的电子通常向更低的振动能级驰豫，在驰豫过程中能量的损失以热释放出。当电子从激发态跳回基态时，经常发光。微阵列中经常使用的荧光团，基态寿命通常仅有几个 ns。

6. 量子效率。荧光团发射光子的概率通常用荧光量子效率或量子效率 Q 来刻画，其定义如下：

$$Q = \frac{\text{发射的光子数目}}{\text{吸收的光子数目}}$$

量子效率在 0 和 100%之间，如 Rhodamine－6－G 的量子效率为 0.9。实际中，量子效率 Q 通常会在其最大吸收波长处列出来，来衡量其发光能力的大小。

7. 荧光团浓度。当激发光波长与强度一定时，荧光团所发射的光子数是荧光团浓度的线性函数。

8. 自身吸收。当荧光团浓度非常高时，由于荧光团分子的密度非常高，激发发射仅在样品表面发生，或发射的荧光再次被其他荧光团分子吸收，此时信号与荧光团浓度变成非线性关

系。

9. 激发光强度。在一定激发范围内，所发射的光子数会随激发光强的增加而增加。如果激发光非常强，大量的荧光分子将被激发到激发态，从而导致光漂白。

10. 环境影响。荧光团所处的环境会影响自身的量子效率、激发光谱和发射光谱。温度、pH 值、离子浓度或溶剂等因素都可能造成荧光团光谱性质和荧光量的变化。

11. 光漂白。另外一个光物理影响是光漂白，它破坏了荧光团的光结构，从而影响染料的荧光强度，这是因为激发态比基态更容易发生化学反应。由于一部分激发态分子参与了化学反应，它改变了荧光团的分子结构，产生了低量子效率或不发光的分子。这种反应的速率依赖与特定荧光团对漂白的灵敏度、化学反应环境、激发光强度和激发光驻留的时间。

12. 抗衰剂。氧分子和第三激发态的染料分子反应经常导致染料漂白，为减少光漂白，可使用抗氧化剂的苯基丙氨酸或叠氮化物、采取无氧条件或降低入射光等来实现。

三、荧光探针的选择

通常探针的选取，应考虑以下几个方面：

1. 激发波长，应根据实验系统所附带的激光器波长加以选择。

2. 发射波长，应根据所拥有的滤光片的波段来选择。

3. 荧光探针的光稳定性和漂白性；在荧光定量和动态荧光监测时，要选择光稳定性较高的荧光探针，同时在实验时减少扫描次数或降低激光强度，来减少光漂白的程度。

4. 荧光的定性或定量。定性研究一般选择单波长激发探针；而定量研究最好选择双波长激发的比率探针，以利于制作标准曲线。

5. 荧光探针的特异性和毒性，根据实验目的，尽可能的选择毒性小，特异性高的探针。

6. 荧光探针适用的 pH。一般情况下，细胞的 pH 处于生理条件下，但是，当探针被用于非生理条件时，应根据 pH 值作出相应选择，同时，配制溶液时也应该注意这个问题。

如图 16－2 是 FITC 激发强度曲线，TITC 是微阵列探测经常用到的荧光探针。从图中可见，在 494 nm 处有一个最大吸收峰值，而峰值两侧曲线很陡。在 518 nm 处有一个最大发射峰。两峰值处波长相差 24 nm，这一现象在微阵列使用的染料中较普遍。

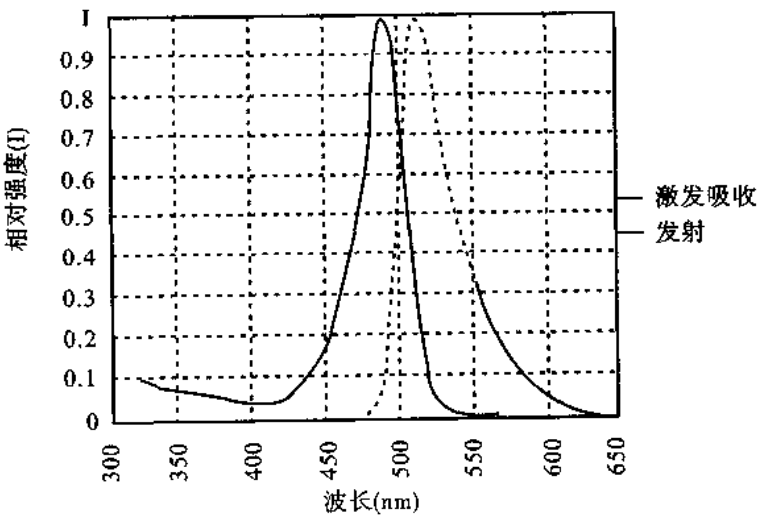


图 16－2　FITC 发射和激发光谱曲线(发射光谱在 400 nm 波长处激发测量)

初看此图，人们可能认为：FITC 可在 510 nm 处激发，会发射 475～510 nm 波长的荧光。事实并非如此，发射波长总是比激发波长要长，激发曲线和发射曲线（如图 16－2）并不是同时绘制的，只是为了便于观察而绘制在同一坐标上。激发曲线的绘制过程是，在单一波长下，变换激发波长在不同波长下测定荧光强度发射，发射荧光曲线的绘制过程则是在最长激发激发波长下开始增加波长，测定不同波长下的荧光发射强度。

微阵列扫描仪设计者，通过研究和分析荧光染料的激发曲线和发射曲线来确定适合某种染料的复合扫描激发波长，该波长的激发效率至少要到达 50%～70%的水平。激发波长要尽量避免靠近发射最大峰值处的波长，否则会干扰荧光信号，因此需要在激发最大峰值左侧选择一激发波长。如对于 FITC，推荐波长在 470～495 nm 之间。

在微阵列探测中，增加激发功率会线性增大发射光的功率（对于单光子激发）。而当激发光强度很大时，发射光强会下降，这时因为荧光输出已饱和或染料被入射光子破坏掉。在微阵列应用中，荧光发射一般比激发光低几个数量级。

微阵列上各个样品点的荧光分子受到激发后，从各个方向发射荧光光子，散射成球状，所以扫描仪应将这些散射光子收集到（图 16－3）。由于扫描仪收集的是很弱的发射光，几何形状是设计扫描仪设备的主要依据。

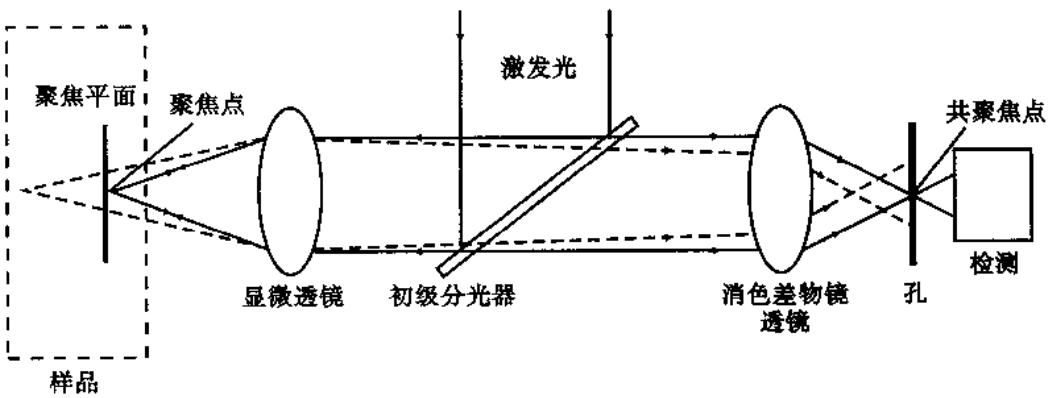


图 16－3　共聚焦微阵列扫描系统

四、探测设备的光学设计要求

微阵列扫描仪有各种形式的，有共聚焦形式的也有非共聚焦的，任何一种设备，如图 16－3 所示，必须拥有下面一些基本功能：

①激发光；②发射光收集；③空间定位；④激发光与发射光分离；⑤检测荧光扫描仪的探测器。

表 16－1　激光扫描共聚焦系统中常用的激光器

波长（nm）	功率（mW）	激光器类型
325	10～30	氦镉（He－Cd）激光
351	300	带紫外的水冷式氩离子激光

续　表

波长(nm)	功率(mW)	激光器类型
364	>20	风冷氩离子激光
442	11	氦镉(He－Cd)激光
457～514.1	25	小型氩离子激光
543.5	1.25	绿氦氖(He－Ne)激光
630	33	外谐振器半导体激光
630	3	碰撞脉冲染料激光
632.8	3.5	氦氖(He－Ne)激光
700～1 100	2000	蓝宝石激光

(一)激发光

激光器、弧光灯、白炽灯或发光二极管(LED)等都可用作激发光源。根据1.2节所讲的荧光知识可知,激发光的波长应限制在一定波长,避免与发射光将发射光覆盖或与之重叠。因此宽带发光光源如白炽灯、发光二极管需要使用滤光片来选择激发光谱。而激光器拥有单色性好、方向性好、功率高等众多优点,用作激发光源不需使用激发滤光片,因此常用作荧光染料的激发光源,表16－1列举了共焦系统中常用的激光光源。白炽灯能发射多种波长的光,使用时只需简单附加一个激发滤光片改变器或其他波长转换器就可提供多种波长的激发。在设备中有时需要使用几种激光器来提供多波段激发波长,这也可用多波长激光器来实现,但是这比多种激光器更难使用和更加昂贵。如图16－4是美国光谱物理公司(Spectra Physics)生产的Ti:Sapphire飞秒激光器,具有连续激发和锁膜激发两种工作方式,它输出波长在650～1 050 nm之间可调。它由Ar^+激光器或全固体的倍频Nd:YAG激光器泵浦。它能满足当前应用的各种功率要求,适合于多种荧光染料多光子激发。Ti:Sappire激光器在～100MHz的重复频率下,能产生100飞秒脉冲宽范围的双光子激发。但是使用和维护很麻烦,相对其他激光器价格非常昂贵。

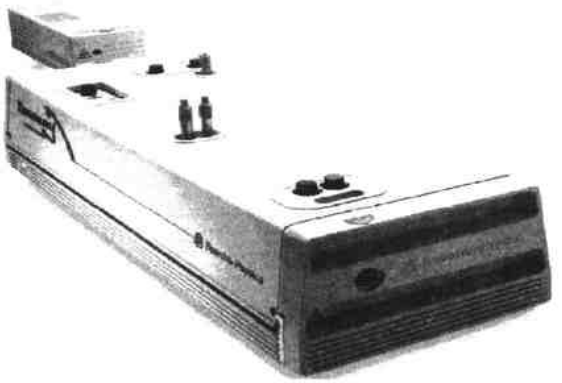


图16－4　Ti:Sapphire飞秒激光器

激发光应直接射向微阵列上的待测样品,通过大量一次照明的方式,待测样品的大部分区

域同时受到激发，这种照射方式在 CCD 数码相机中较常见，而很少在共聚焦扫描仪使用。CCD 探测器热电制冷通过减少电子噪声能巨大提高其灵敏度，因而产生近似为零的背景，具有高动态范围（16 位 = 216 灰度级）以及高量子产率。使用这种方法时，样品激发区上照射不均匀是仪器设计者应当考虑的主要因素之一。

另一种可供选择的办法是，激发光被会聚成很小一个斑点来照射样品上的一小部分。这种方法采用特定的光线收集和定位，当聚焦光扫描样品时，在极短时间内、在给定的区域上会聚光束提供很强的激发，通常可达到 10 000 W/cm^2。

依据第一节第二点所讲述的，根据所使用的染料特征来选择激发波长。激发波长应当在染料激发峰值的左侧，并且激发效率很高的区域。激发效率越低，就需要提高激发光功率以提供一定水平的荧光。激发光过强可能会产生光漂白，损坏样品或污染荧光发射信号，降低荧光信号。

（二）发射光收集

经常用物镜来收集荧光，它是一个透镜，把激发光会聚到样品上，使在一定角度范围或锥形体内的发射光到达探测通道。光收集角度区域大小是很重要的参数。当荧光发射是球形时，光的收集范围是决定仪器收集效率的第一决定因数。透镜的光收集角度经常用数值孔径 NA 来刻画，它表征了透镜收集效率的大小。数值孔径是物镜三个参数中最重要的参数，常用下面式子来表示：

$NA = n \cdot \sin(\mu/2)$

式中：

n - 为物体与物镜之间介质的折射率；μ - 是进入物镜锥形光柱的角度。

数值孔径也决定了物镜的分辨率，越大物镜分辨率越高，但是会使物镜工作距离下降。图 16-5 描绘了数值孔径与收集效率的关系曲线，它的发射源是一个球状点源。

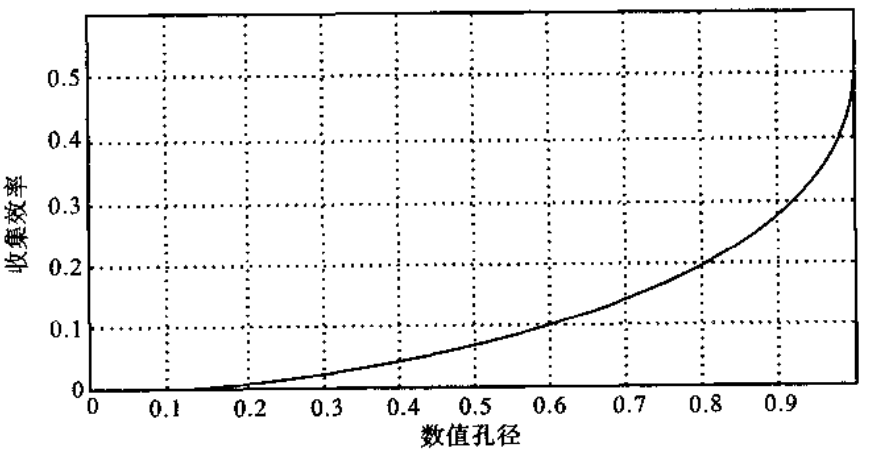


图 16-5　不同数值孔径下物镜收集效率曲线

数值孔径为 1 的透镜表示其能对整个半球区域收集光，相应的收集效率为 50%。在许多共聚焦微阵列扫描仪的数值孔径在 0.5 ~ 0.9 之间，而许多以 CCD 相机为基础的阵列扫描仪的数值孔径在 0.2 ~ 0.5 之间。

物镜另外两个参数是放大倍数和物镜的制作材料与工艺，放大倍数是 3 个参数中最不重

要的，它只决定对物放大多少倍，而不决定分辨率和亮度。而物镜的制作材料具有提高物镜成像质量及消除其他光学瑕疵等功能。

一些阵列扫描仪并不使用物镜，而使用一个积分球面镜来收集部分发射光，但是部分光线经过多次球而反射而衰减。还有一些无散射光收集装置，仅仅是将探测器放在样品的某一区域的上方，光线收集的效率受限于样品被照明处的范围及探测器的视角，例如，一个直径为 25 mm 的探测器放在激发点上方 100 mm 处，其实际的有效数值孔径为 0.12 mm。

(三)空间定位(Spatial addressing)

空间定位指对样品上某一小区域上的荧光信号进行荧光强度计算，微阵列上各点样品被分割为许多微小的像素，在进行荧光信号定量时，空间分辨率必须高于样品点的大小，如微阵列上样品点的直径为 100 μm，则像素的大小为 5 ~ 20 μm。

使用多元探测器阵列如 CCD 或使用机械扫描装置可实现空间定位。许多 CCD 相机被设置为大范围照射，探测器直接提供许多像素组成的图像，该方法的缺陷是 CCD 探测器所提供的数值孔径较小，且需要后向照射和维持系统冷却设备价值昂贵，由于光散射会导致像素之间交叉重叠造成信号不清晰。

而机械扫描方法，需要设计把激发光束会聚到一个像素大小的点上，然后使用一个单元探测器从这微小的点区域收集发射光。为了覆盖整个样品区域或为了对整个样品区域成像，进行移动或扫描是必须的。通过镜子扫描光束、或移动样品、或两者的结合来实现扫描功能。与 CCD 成像系统相比，扫描会给系统增加更复杂的机械装置，但这种方法可使用高数值孔径的光学收集系统、使用不太昂贵的探测器，同时可获得更高空间选择性。当样品信号很弱时，更高光学收集效率是必须考虑的。

(四)激发光与发射光分离

微阵列上各点的荧光强度通常比激发光功率低几个数量级，要从激发光中检测出微弱的荧光信号，就需要在荧光信号探测之前采取一些光学措施来分离这两种类型的光束。尽管特殊几何设计可以分离两种光束，但由于激发光和荧光的波长差别，波长分离应成为首选的方案。

许多使用物镜的微阵列扫描仪采用的是表面照射方式(epi - illumination)方式，激发光和发射光沿着相同的路径从物镜到样品，但二者方向相反。这种设计就会使从样品上反射和散射的光与荧光混合在一起。所以需要分光镜来分离出混合光，它也是光束分离的第一道分离装置。

其中一种类型的光束分光镜是分离颜色的二色滤光片或多色干涉滤光片，它将激发光反射而让波长更长的光透射过去。图 16 - 6 是 OCLI 公司生产的高性能二色镜透过率的光谱曲线。图 a 是红光波段的二色镜，它只能让波长 580 nm 以上的红光通过，而低于 580 nm 的可见光被反射。图 b 是绿色二色镜，它只能让波长 490 ~ 580 nm 之间的绿光透光，其他波段的可见光被反射。这种分离器经常使用，对 1 ~ 3 种不同激发与发射光能进行很好的分离。但若超过四种以上的混合光则分离很困难。

另一类型的光束分离器是几何光线分离器，如图 16 - 7，它采用几何的方法来分离激发光与发射光。在扫描系统中，物镜数值孔径为 0.6，像素大小近似为 10 μm，从透镜出来的激发光束直径比发射光细。一块小反光镜将激发光反射，而让环形部分发射光通过。作为与波长无关的分离器，其分离效果与波长分离器相同。

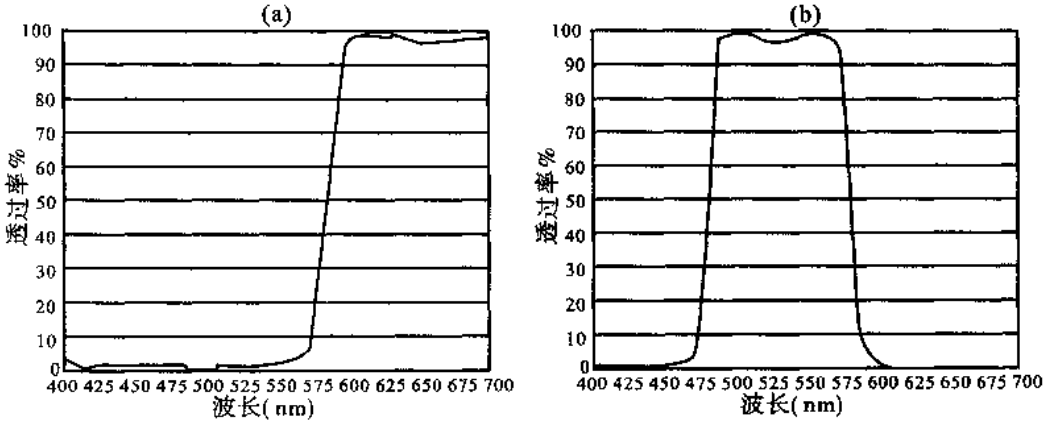


图 16-6　二色镜光谱曲线(横轴是波长,单位是 nm;纵轴是透过率)

(a) 红光二色滤光片;(b)绿光二色滤光片

理论上来讲,性能良好的分离器应能完全分离出激发光和发射光,但事实上并非如此。通常还要在探测器之前放置发射滤光片来过滤发射光,它只允许染料发射峰值附近很窄的一段波长的光通过,而其他光被阻挡掉。这是微阵列扫描仪必须的第二道光束分离装置。有的扫描仪不用光束分离器,而是让激发光和发射光在不同轴向上。该方法能将发射通道上的激发光反射回去,但却很难达到较高的数值孔径。因为透镜距离样品很近,通常小于 1 mm,所以激发光能进入透镜的角度范围很小。

其他器件也可用于波长分离,如三棱镜或光栅。这些器件拥有一些令人感兴趣的特征,如连续的光波调协功能。但是在微阵列扫描中,由于要求对激发光有高度的敏感性,因而对发射装置也相应要求有很高的精密度。

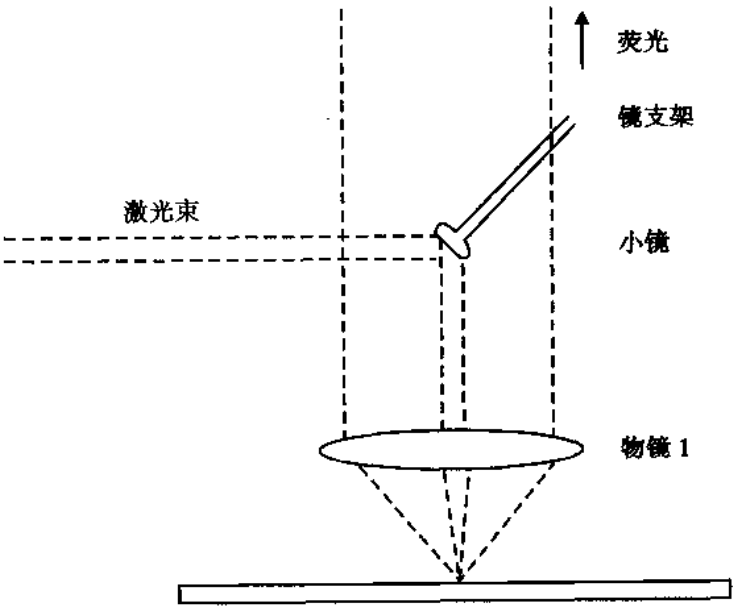


图 16-7　几何分光镜光路结构图

(五)检测荧光扫描仪的探测器

微阵列扫描仪的探测器能把微弱荧光信号转化为电信号,在阵列扫描器中常用的探测器主要有光电倍增管(PMT)、CCD 阵列和雪崩二极管。每种探测器都有自己的优势也有自己的缺点(它们之间的差异在第三节里详细讲述),应根据应用的特定条件来选择合适的探测器。

PMT(photomultiplier tube)是可见光范围内最灵敏的探测器,它属于单点探测器,需要扫描系统来对微阵列上的样品进行空间定位。改变 PMT 的电压可以很方便地改变光电倍增管的灵敏度,而其电压调节可以满足阵列扫描器所需的灵敏度要求。PMT 灵敏度在红外和近红外波段会急剧下降,因此它们的使用通常限于可见光范围。如图 16-8 所示,是日本滨松公司生产的一种光电倍增管光谱响应曲线,它表征了光电阴极灵敏度与波长之间的关系。从图中曲线可观察到,该管子对可见光范围的光响应灵敏度高而且量子效率高,尤其对蓝光和青光最灵敏,而在红外波段灵敏度和量子效率急剧下降。

CCD 阵列不具备 PMT 自身拥有的内在低噪声放大功能,因此需要外部放大器来达到与 PMT 一样的最大灵敏度。其主要缺点是,CCD 探测器的实际数值孔径很难达到 0.6~0.9 的范围,这限制了 CCD 阵列可探测信号的范围。如在弱光条件下,对这要求非常高,因为光子流的统计特性完全决定仪器的探测底线。另一个缺点是 CCD 不能与共聚焦系统相兼容。当入射光很强时,因为 CCD 探测阵列有内置式扫描功能,其优越性就体现出来。

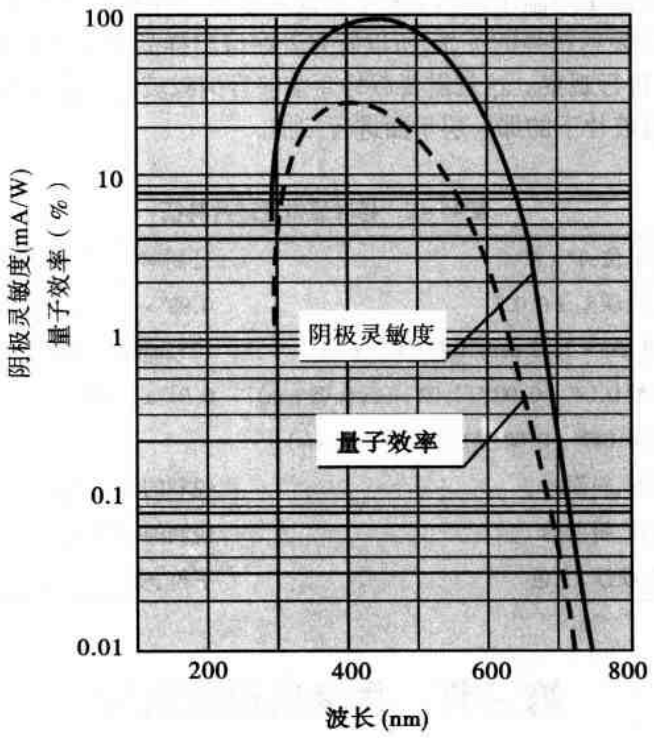


图 16-8　PMT 探测器阴极灵敏度及量子效率

第二节　样品处理

所有微阵列扫描仪都有固定和放置微阵列固相基质的装置,我们称之为扫描仪的载样盒。

载样盒要有能够容纳下玻片的空间，组成其的材料应能经受住上千次取放样品的刮擦考验。通常微阵列的固体介质为显微载玻片，还有塑料片，但是塑料片不如玻璃片刚硬，容易变形，不易聚焦准确，所以多数研究者使用的是显微载玻片，因此本节主要讲述用于微阵列探测中显微载玻片的规格及相关的处理。

一、标准显微载玻片

事实上显微载玻片已成为微阵列的标准载玻片，占据着市场的主导，但是并没有唯一的标准规定显微镜载玻片的规格。它们广泛被做成各种尺寸、厚度、楔形和拐角构造以及进行末端处理。表 16－2 总结了许多美国实验室供应商提供的载玻片，它们具有不同尺寸和偏差。标准的扫描设备应当能够容纳多种规格的载玻片。

二、其他规格载玻片

微阵列固相介质还使用其他格式的玻片，如塑料制成的载玻片、具有各种形状和尺寸并充满液体细胞的专门设计的玻璃载玻片，每一种载玻片都采用特定样品固定方案。例如塑料没玻璃坚硬，在共聚焦扫描器中更难保持平滑和聚集。

三、样品环境

扫描检测时通常是在室温下进行，此时载玻片上各点样品已经干燥，然而，有些染料如 FITC 在潮湿环境才能发射很强的荧光，所以研究者在待测样品上滴加适量的缓冲液把阵列弄湿，然后盖上盖玻片进行观察。这就需要载样盒能够容纳载玻片与盖玻片叠加之后的厚度，扫描仪在扫描时聚焦盖玻片下的那一层平面进行扫描。

表 16－2　标准显微载玻片特征

盖玻片类型	英寸单位：1″×3″	公制单位：25 mm×75 mm
X－Y 大小	1″×3″，误差 ±0.02″	0.98″×2.95″，误差 ±0.02″
	(25.4 mm×76.2 mm，误差 ±0.5 mm)	(25 mm×76 mm，误差 ±0.5 mm)
厚度	标准型 0.04″±0.002″(1.02 mm±0.05 mm)	0.04″±0.002″
	厚型：0.047″±0.004″(1.2 mm±0.1 mm)	
拐角	锐利的、斜面的	锐利的、斜面的
边缘	锐利的、斜面的	锐利的、斜面的
末端表面处理	平整、喷沙，上色	平整、喷沙，上色

第三节　共聚焦扫描原理

前面几节介绍了微阵列相关的一些知识，本节将讨论共聚焦原理、光学系统设计以及如何工作的，并归纳出共聚焦系统的优点和缺点。

一、共聚焦原理

在前面各节知识基础上，这里将详细说明共聚焦原理。正如其名字所暗示的，共聚焦扫描

有两个共焦点，这两个焦点限制了三维视场的大小。这种设置仅能收集焦点内的微小区域，而焦点以外微小区域的光被阻挡掉，同时也滤去了许多图像瑕疵。共聚焦系统仅仅对一个小区域（可能是一个点或一个像素）成像，因此需要扫描才能获得二维多点图像。这里以反射式荧光扫描共聚焦显微镜为例，说明扫描共聚焦显微镜的基本原理，然后介绍共焦微阵列扫描仪的特征。

如图 16－9 所示，激发光源经过照明针孔形成点光源，通过激发滤光片后被二色镜反射到显微镜的物镜上，被物镜聚焦到 3D 样品内。物镜在样品区域会聚，形成的焦点区域称之为焦平面，在焦平面和焦平面上下区域被激光照射而激发出荧光，所发出的荧光又被物镜收集，通过二色镜和发射滤光片，到达共焦针孔。由于照明针孔和探测针孔相对于物镜焦平面是共轭的，只有焦平面发出的荧光（我们以黑线—表示）能在共焦针孔处会聚一点，并通过针孔到达探测器（PMT 或 CCD），焦平面外（我们以－－－－表示）的点所发出的光被共焦针孔阻挡掉，不能被探测器探测到。所以焦平面以外发出的光都被针孔挡住了，对共焦图像贡献很小。这样只能得到样品区域在物镜焦平面某一点的图像，而要获得样品的光学横截面（被焦平面切割的断层）的二维图像，则需要在垂直于光轴的 xy 平面（焦平面）上进行扫描才能获样品二维图像。如果逐步调节样品 Z 轴（光轴）位置，可产生样品的多幅断层图像，细胞或组织各个横截面的图像都能清楚的显示。这种成像方法称为共焦扫描显微镜层析成像方法，通过三维重建技术，则能得到样品的三维立体图像。如果在成像系统中移去共焦针孔，那么焦点外的光线也到达探测器，深度方向（Z 轴向）的分辨率就大大降低了，那么非共焦的荧光显微镜就和普通的显微镜没有多大差别。

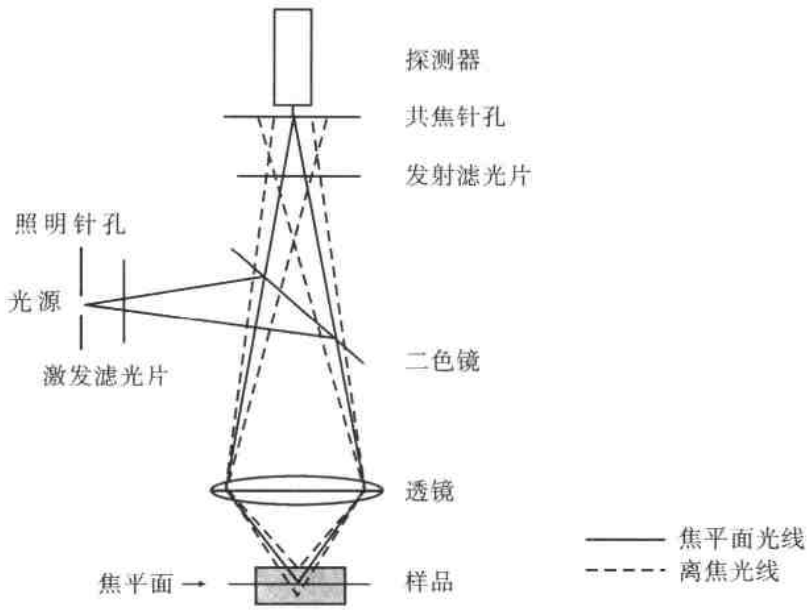


图 16－9　反射式荧光显微镜共焦原理

二、共焦微阵列扫描仪

本节以实际光路图，详细讲述微阵列焦点处探测和焦点外探测特点，讨论共聚焦扫描系统

设计的要求，在此基础上归纳出共聚焦扫描仪主要优点。如图 16－10 所示，平行的激光被分光镜反射到物镜上，激发光束仅仅填充物镜一部分。填充量额的多少取决于所选择物镜的数值孔径和所期望像素的大小。激光被物镜会聚到样品上，在那里激发出的球面荧光向各个方向传播。激发光也被反射朝着探测器方向传播。

物镜收集部分球面荧光并校正使其成为平行光。物镜也收集被反射的激光，它比荧光强度高 3～7 数量级。返回的光束直接到达分光镜，分光镜把多数激光反射回激发源处，而允许大部分荧光通过分光镜。一块平面镜在不经任何光学处理将荧光反射到发射滤光片处。由于发射滤光片只允许窄范围波长的光通过，阻挡掉所用剩余的激发光。

从上面共焦原理可知，共聚焦的本性就是拥有探测器物镜和共焦针孔。物镜把平行光会聚成直径很小的光束，共焦针孔阻挡焦点外的光而让焦点处的光通过。这就限制了物镜聚焦的深度，图 16－11 显示了焦点外光路通道，发光点不在物镜焦点范围内，则光线将被探测物镜聚焦于共焦针孔前面而被阻挡掉，只有焦点范围内的光线才能有效的通过共焦针孔被探测器检测到。

比如一块灰尘在样品焦点下面形成第二点光源，它所发出的光也被物镜收集。由于第二点并不在物镜的焦点处，沿探测器方向前进的荧光光束并不能被校正成平行光，而会相交一点，在针孔前面而不是在针孔处形成一个焦点。在焦点之后光束扩展，形成很大直径光束打在共焦针孔上。从而阻挡掉许多焦点外的光。在物镜焦点上面的灰尘，共焦针孔具有与下面相同的阻挡效果，它只能让焦点处很窄范围处的荧光能通过共焦针孔，因而，共聚焦扫描仪获得图像的信噪比要高于非共聚焦扫描系统。

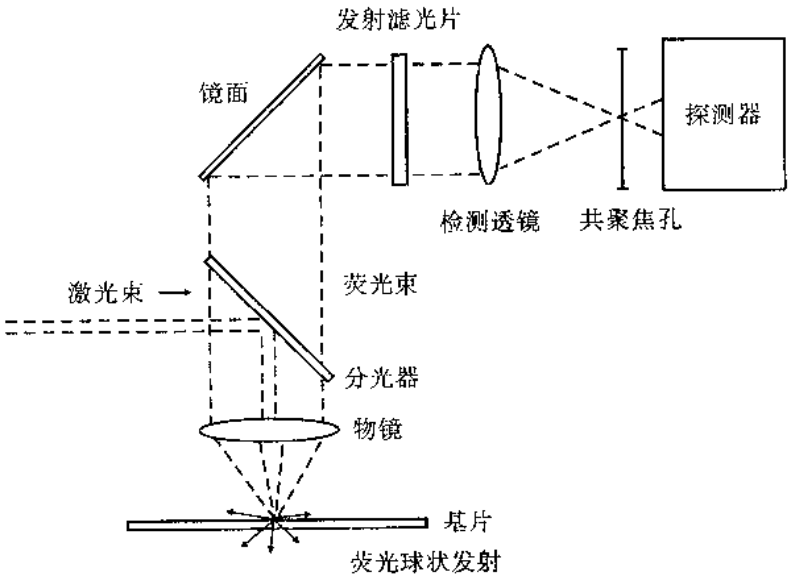


图 16－10　激光共聚焦微阵列扫描系统

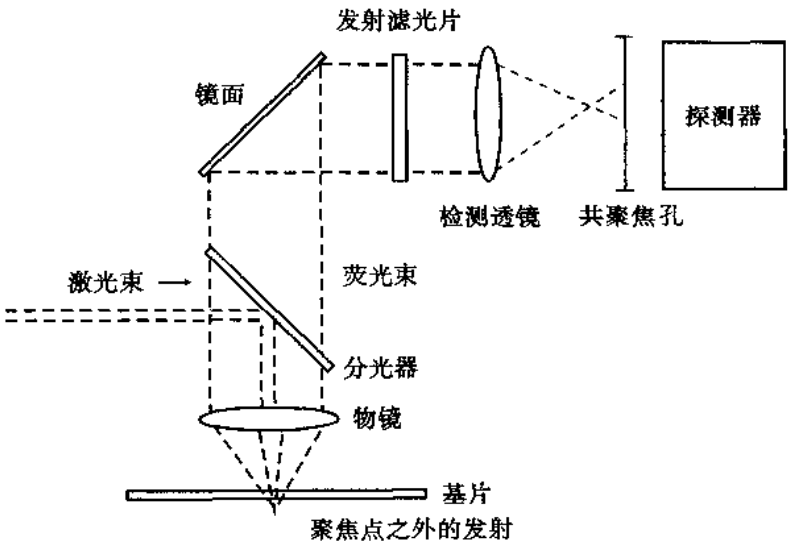


图 16－11　激光共聚焦系统焦点外光路图

三、共聚焦系统优点与缺点

共聚焦生物芯片扫描仪的主要优点：

1.生物芯片上的照明光仅仅是扫描中非常小的一个聚焦光点，因此亮度和信噪比高。

2.探测器(如光电倍增管)只能接收通过共焦针孔的光，而来自生物芯片其他部位的杂散光因在共焦针孔处不能聚焦而被滤除，结果得到的是一个高分辨力和高对比度的图像。

3.由于成像光束的聚焦与波长有关，在光路中加入窄带干涉滤光片，从而进一步提高了图像的清晰度。

4.共焦扫描成像系统采用轴上光束，因而像面没有像差变化而产生的像畸变。

5.因采用物体扫描，光路稳定，充分发挥了光学系统的中心优势。

6.可进行三维扫描成像，这是一种特殊的光学层析成像方法，完全不同于一般计算机层析成像方法。它用计算机控制对样品进行分断层、逐行、逐点完成三维扫描，同时收集图像数据存入计算机存储系统。根据需要由计算机图像系统显示任选的二维断层图像或三维透视图。

7.由于激光的亮度高、单色性好、发射角小、聚焦性优良，因面在生物芯片扫描仪中采用激光作为光源，成像系统色差小。因为高数值孔径、高放大倍数的物镜均不可避免地存在色差，用激光光源可以提高光学系统的分辨能力。

但另一方面，对聚焦范围的严格限制要求载玻片摆放非常平稳，扫描运动过程中微阵列要保持在同一水平面上，偏差要小于焦距的范围，偏差小于 ± 10 μm。这一要求增加了机械加工的精密度，使得加工工艺更加复杂，但这是保证获得最佳图像质量的有效途径之一。在各种系统扫描光束或移动载玻片时，载玻片一定保证平坦，并且它的运动也得保持平稳。这就增加了机械和光学设计的难度，但这能极大提高图像质量。

四、共聚扫描实现与折衷考虑

由于激光聚焦于载玻片一点上，要获得整张玻片上各点的荧光信息则要对玻片各点进行扫描成像。可通过移动扫描装置或移动载玻片来实现上述功能。正如图 16－10 所示，在分光

镜和物镜之间插入一个扫描装置，使物镜能在平坦平面内扫描，就可实现移动光束扫描功能。扫描装置移动后要求物镜能采集到荧光，这样的物镜很复杂，一般不超过 0.3 或更小。

较为简单的且高收集光效率的办法是移动物镜或移动载玻片。能够移动载玻片的设备在平稳光学系统上以光栅的方式移动载玻片。但移动会影响扫描速度。不过在弱光条件下，图像读取速度取决于探测到光子的累积量，所以即使快扫描速度和低光学收集效率的话，图像读取信号也未必快。

第四节　共聚焦扫描仪光学设计及关键特征

在了解共聚焦微阵列扫描仪共聚焦工作原理的基础上，本节将详细探讨建立一套共聚焦微阵列扫描仪系统所要考虑的关键技术，并详细说明了扫描仪各部件的功能参数。

一、波长鉴别

所有扫描仪都会面临同样一个问题，即如何将背景降低、提高待测样品的信号。共聚焦技术的应用使得共聚焦扫描仪能比其他类型的扫描仪获得更高质量的图像信号，但使用该技术仍不可避免地收集到激发光和其他杂散光，如何从背景中分离出荧光信号，是检验扫描仪可用性的关键。这里以此为出发点，详细说明波长分离的折衷考虑与设计，即在探测和定量评价信号之前从其他光中分离出荧光。

微阵列扫描仪可以探测到 pW(10^{-12} W)级的光信号。许多材料都自发荧光，如玻璃载玻片、附在载玻片上的化学物质、样品冲洗遗留下的化学物质、透镜、滤光片甚至 DNA 样品自身也会发荧光。为减少其他光信号探测，需最佳化探测系统，以收集到有用信号。另外，反射或散射激发光可能比荧光信号强一百万倍，如何抑制掉它也是要着重考虑的，如果不能从其他光束中分离出荧光信号，就会造成很高的背景信号。而背景信号会决定染料浓度探测的底线，因此要尽可能减少背景信号。

(一)分光镜

不同共聚焦扫描仪采用不同类型的分光镜，如图 16－10 所示，分光镜需要从激发光中分离出荧光。玻璃载玻片的第一层反射大约 4%的激光，一些研究者正研究使用镜面反射载玻片以使更多荧光返回到探测光路中，一块镜面反射载玻片几乎反射所有光束。

分光镜应能在探测通路上阻止许多反射的激光，而让多数荧光通过。常用的器件就是二色镜和多色干涉滤光片，它们通常对一种波长、两种波长、或三种波长激光达到 90%的反射率，而对某一荧光、两种荧光或三种荧光的透过率是 60%～85%。尽管有 90%的激光被消除，但是仍有 10%的激光会到达探测通道，它的过滤就要求助于发射滤光片。这些分光镜工作的很出色，但要应用到更强的激光或更多的荧光波长中变得较困难。

几何分光镜如图 16－7 所示，比二色镜明显能提高分离性能，且对波长没有依赖性。它的应用要考虑众多光学参数，以便协调更好的工作。使用这种分离器的首要前提是激发光束一定比返回发射光束小的多。通过选择会聚激发光尺寸、焦距长度和物镜的数值孔径可达到此目的，通常该技术比高数值孔径方式工作的好。使用一个数值孔径为 0.75、焦距为 6.6 mm 的物镜，若输入激光直径为 0.75 mm，物镜小孔为 10 mm，则会聚的激发光半径大约 600 nm 左右。这种技术工作的很好，能传递 80%的发射光，同时差不多阻挡掉所有反射的激光。已证实这种分离器能在非常微弱样品信号中提高信号/背景之比，比二色镜提高 1～3 倍。其他类型分

光镜，如极化分光镜和 50/50 带通分离器在微阵列扫描器中并不能充分阻挡激发光。

（二）发射滤光片

共聚焦微阵列扫描仪中的分光镜自身并不能完全从其他光信号中分离出荧光信号。如图 16－10 所示，在共聚焦探测器之前，需在校正光路上放置一个发射滤光片。发射滤光片应能让荧光信号通过而阻挡其他光，图 16－12 是一个低通发射滤光片，它只能让波长低于 575 nm 的光通过，而阻挡其他波长的光通过。

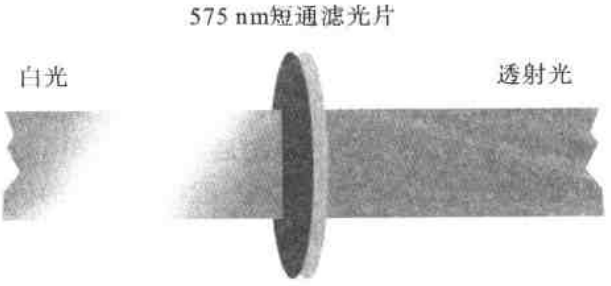


图 16－12　低通滤光片工作原理

从图 16－13 中 FITC 的荧光发射曲线中，我们能推断最好的发射滤光片应是包含所有发射曲线的带通滤光片，如选择波长 510～600 nm 的带通滤光片，它会尽可能多的把荧光传递到探测器上。事实上发射光是很复杂的混合光，除了染料的发射曲线外，还伴随其他发射光曲线，如玻璃载玻片的荧光，载玻片表面处理后留下的化学物质和冲洗样品遗留下的化学物质发出的光信号等。这些光的荧光曲线通常难确定而且是经常变化的。当降低这些干扰荧光时，发射滤光片需要最佳化荧光信号，实现此目的的最好方法是利用一个窄带通滤光片，其带通中心波长与染料发射峰值重合，如图 16－13 所示。该图显示一个带通滤光片，在光强一半处它的带宽是 10nm。它能让很窄的窗口范围内的荧光通过，而阻挡带通以外的光，从而提高信噪比。此方法的不足之处就是降低了信号的输出量。实验证明 10nm 带宽滤光片能提供良好信噪比，它在峰值处有 60%的透过，如图 16－13 所示，仅有部分染料发射光谱与带通窗口重合，这样一些滤光片通常仅达到 10%～15%的透过率。

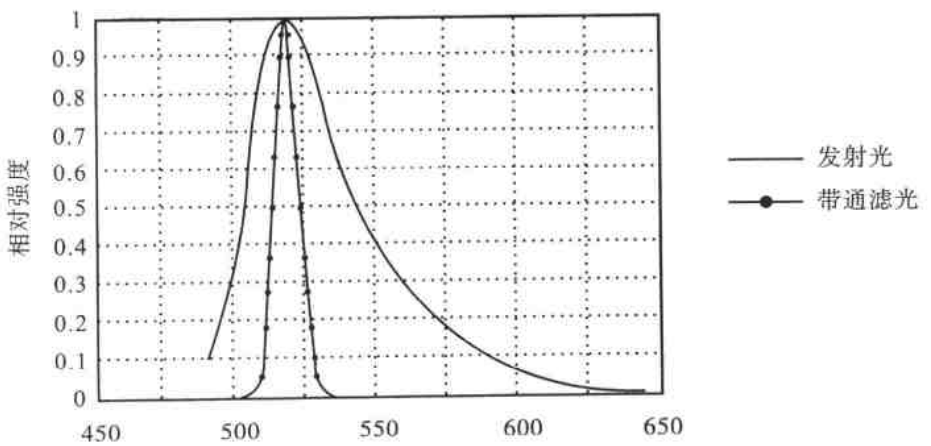


图 16－13　10 nm 带通滤光片光谱与 FITC 发射光谱

窄带滤光片在阻挡其他干扰光的同时也降低了入射荧光信号，因而它仅仅用在高收集效率的设备中，如高数值孔径的物镜中。使用带宽 10 nm、中心波长在染料发射峰值处的滤光片，用数值孔径为 0.7 物镜的微阵列扫描仪工作得很好。更宽的滤光片会使背景光比信号更多的透过，而窄带滤光片能使信号和噪声一起下降。

二、探　测　器

激光扫描共聚探测系统中使用较多的探测器是 PMT 或 CCD，它们把接收到的光电信号转换成电信号，而后被采集卡送到计算机内进行图像处理。下面具体介绍 PMT 和 CCD 的工作特性。

PMT 的量子效率一般较低，所以在 PMT 的生产工艺上尽量减少透射光和反射光的损失，如以不同的介质制作 PMT 窗，使光以一定角度入射后在窗内部产生 2 次反射回到光阴极层，就可以显著改善 PMT 的量子效率。另一方面，现在一般都将 PMT 的信号转换成数字信号来处理，可提高量子效率。虽然 PMT 倍增噪声可能较高，但通过后级的脉冲计数技术或相对增加计算时间便可抑制噪声。所以 PMT 是在激光扫描共聚焦微阵列探测中应用最为普遍的探测器。PMT 的缺点是受温度影响大，但在激光共聚焦微阵列扫描仪中的荧光信号幅度的变化还不致于引起 PMT 温度的较大变化。

CCD 量子效率可高达 10% ~ 80%，在一些特别波段，甚至可达 10% ~ 100%，灵敏度极高，动态范围大，可达 10^5 ~ 10^6级，适合激光共聚焦微阵列探测的信号亮度范围。线性度高，为后续的图像处理提供了最优条件。结合双向扫描，使用 CCD 还确保了图像的高几何保真度。但它的缺点是噪声较大，要使其噪声下降到可接受水平，则需要很低的温度，所以现在一般都要使用制冷 CCD。

传统硅探测器如 PIN 管不具备像 PMT 那样的低噪声增益，而且在可见光范围有非常低的量子效率，因而它们需要外部放大几百万倍以达到光信号灵敏度。外部的放大会增加噪声，从而降低信噪比。硅探测器对波长有一个响应范围，在波长 800 ~ 900 nm 内达到最大，因此它们在近红外具有更大的优势，而 PMT 通常在 500 nm 处达到最大，在 650 nm 以上响应会急剧下降。

探测器的性能与信号质量有直接关系，如在共聚焦微阵列扫描仪中使用 PMT 探测，有许多因素会直接影响 PMT 的输出信号。比如：激光功率、温度效应、探测器增益、稳压电路的稳定度、透射的特殊波长及所用滤光片、物镜的数值孔径及一些光学器件的透射等等。所以，微阵列扫描仪能否获得好的数据，还需要众多装置的协调与相互配合。

三、信号处理及仪器控制

任何微阵列扫描仪中都涉及样品处理、光学系统和探测设计这些核心，对于广大生物学研究者而言，不讲述信号处理及扫描仪控制方法是不完整的，这里将讲述整套整设中需要考虑的控制及处理功能。

（一）灵敏度范围

在微阵列样品制备过程中，步骤较多，如 PCR 反应、杂交反应条件变化度、试剂存储条件不同、不同类型的基因表达等因素的影响，即使使用相同的染料，其亮度也可能相差 1 000 倍，在同一块微阵列上的荧光强度范围达到 4 000:1 的范围。大部分扫描仪的荧光强度读取范围在 4 000 ~ 16 000 之间，若不改变仪器的灵敏度范围，就没有足够的动态空间来容纳这种样品

的变化、无法适应于各种荧光强度样品的读取。

调节仪器灵敏度范围有两种方法：一种是通过调节激光衰减器来改变激发功率，激发光通常变化范围至少是 100:1。第二种方法(对于使用 PMT 作为探测器的微阵列扫描仪而言)是通过调节电压输入来改变 PMT 的灵敏度，其调节范围也至少是 100:1。这两种方法可相互叠加使得仪器的调整范围是 10 000:1，这一范围对普通的扫描仪都足够了。

通常人们希望使用的激光强度大而不会破坏样品上的荧光分子，我们已经知道，激发光越强越大就会激发荧光光子越多，信号也越清楚和强烈。如果样品经过多次扫描后，需要降低激光功率以缓解累积光子对样品的损害。

由于这些仪器有很宽的调节范围，需要多次实验、操作才能找到恰当的调节结合点。幸运的是，许多商业化的共聚焦扫描仪有自带的功能，可以自动设置灵敏度以获得高质量图像，不仅缩短了仪器工作时间，也为用户节省了许多时间。

(二)信号转换及采样过程

信号转换过程就是把受激发的荧光染料所产生的光信号通过一系列过程转化成连续的数字信号。荧光信号是由一系列随机光子发射累加而成，若将它们累计可与某一样品区域的染料分子的密度相关。通常扫描仪对空间上和时间上探测到的荧光光子的数量都加以累计并取平均值才能代表某一样品区域染料分子的密度。

扫描区域被分成尺寸相同的像素，激光共聚焦和其他点照明式扫描仪对逐个像素激发，然后探测器累加各像素点上荧光光子。激光经过物镜的会聚，在样品上形成一个斑点，成为激光斑。它扮演一个能移动空间位置、平均处理数据的机械装置的角色，从整个斑点区域内发射的光在探测期间被相加。

激光斑的尺寸与形状决定了其自身的平均特征。激光光束服从高斯强度分布轮廓，这种构造导致了斑点中心处的染料分子更容易被激发，能产生更多的光子，因此在平均处理过程中斑点更会加大比重。通常可选择激光腰束半径等于像素大小的激光，更大的斑点将会把更多像素平均起来，较小斑点可能包含不到空间平均过程中完整的像素。

光信号被 PMT 探测到并被转化为电信号，在转化为数字信号之前需要进一步处理。电信号仍旧含有与光子发射关联的随机过程，数字化处理过程相对像素驻留时间是瞬间的。使用电路滤波把光子发射信号平均为一个信号，该信号就代表了每个像素包含的信息。

在快速数字化转换之前，可用指数衰减的低通滤波器来进行时间平均。当对一个分段函数滤波时，这些滤波器能产生一个指数衰减的信号。滤波器的时间常数是进行恰当的数据转换参数，时间常数就是幅度降为初值 1/3 所花费的时间。实际中，时间常数被设置为像素周期的 1/3 或 2/3。更长的时间常数会把像素平均在一起，而较短的时间常数可得到单个光子事件，这会使最终数据失真。

(三)暗噪声信号

即使没有光存在，探测单元也会产生一个非零信号，这称为暗噪声。一定程度上暗噪声在一次成像扫描中恒定的，可以从信号中减去，从而并不影响仪器的动态范围。

(四)图像获取、显示和存储

共聚焦微阵列扫描仪需要对样品进行扫描，以便获取荧光强度分布的数字图像。无论是移动光束或移动样品或是二者结合的扫描系统一定对样品区域进行光栅式扫描。为满足基本要求，扫描区域应当可调，以便当所感兴趣区域比样品小时，用户不需等待仪器扫描整个样品

结束，就可调节扫描区域，从而节省更多时间。另外，低分辨率、高速扫描也是受欢迎的。如在很大视场内，可以非常迅速的扫描5条线或一小块区域。

各像素的分辨率可达到16比特，16位是一种很方便的图像格式，方便于后续的处理和操纵。当正在扫描图像时，显示器要显示图像，以供扫描者确定样品选取是否正确、仪器设置是否正确等等。经常把荧光图像转化为伪彩色来显示，便于人们在显示器上观察。

扫描完图像后，可将图像数据保存为图像格式文件，如tiff格式，这种格式的图像文件是目前用途较广的一种图像文件。

(五)控制和自动化系统

控制和自动化系统由电脑来完成，扫描仪使用者无须精通扫描仪的光学和电子学部分，已商品化的扫描仪已经过大量的测试和信息反馈并不断改进，使用户界面越来越友好，图形化的操作界面使得初学者只需要接受简单的训练或阅读操作手册就能快速使用仪器。控制软件具有许多自动设置功能，让使用者节省时间、提高工作效率。如不同的用户扫描不同样品可设置、保存、再现扫描参数；可自动设定两个以上的扫描激光波长并自动保存图像结果；对某一扫描波段自动调整灵敏度、激发光强度及探测器探测范围。

扫描仪或它的主机也应提供几十亿字节的存储、网络联接和支持大容量远程存储。

四、扫描仪功能参数

微阵列扫描仪是复杂的仪器，拥有大量参数和规格。这里罗列出最重要参数及其性能测量，它决定了仪器在应用中是否合适。

(一)激光源数目和荧光波段

激光源数目是扫描仪的基本参数，单一激光器只能激发几种染料和使用几种规格的发射滤光片。第一代扫描仪带有2个激光器和拥有2个或4个荧光通道，而第二代仪器则含有3到4个激光器，其荧光选择通道可达到10个。

(二)探测能力

探测能力就是设置灵敏度，使最亮点达到光强满刻度时区别于背景最小样品点的荧光亮度。其单位是单位面积上染料分子的数目(fluors/ μm^2)。因样品制备过程具有不确定性，生物学研究者通常并不知道染料密度是多少，通过它的测量可判定染料的浓度。探测能力通常用特殊制作的微阵列来标定，它通过标准样中样品点的探测来确定。

样品或仪器决定了探测能力的极限。如果样品制备过程中，造成样品点之间很强的背景荧光，样品就会受到限制。如果仪器引起的噪声不能减去，并且超过样品背景水平，则仪器的使用将受到限制。

(三)灵敏度

灵敏度是在特定波长处仪器能探测的最小光强，它是对仪器获取最弱信号的衡量。与探测能力不同，灵敏度并不依赖样品的特性。当把激发条件和探测条件都调到最大值时，如果对样品的扫描速度很低时，灵敏度就会限制探测能力。

(四)干扰

扫描多种染料标记的样品时会发生干扰，干扰就是染料的激发和探测发生错误或使用不当的激发波长和发射滤光片。使用中心波长在染料峰值处、对带宽以外波长具有良好衰减的

窄带发射滤光片可降低这种干扰。通常选择染料的最大发射峰值处的波长与激发光波长相差很远，至少 50 nm 以上，可大大降低这种干扰。

（五）分辨率

微阵列扫描仪的空间分辨率通常用像素大小来表征，许多商用设备分辨率是 5 μm、10 μm 或 20 μm。

（六）扫描视野

视野大小就是载玻片上被扫描的面积，它应与阵列制备过程相匹配。扫描区域越大，每个样品应放置更多的点。通常载玻片周围 1 到 1.5 mm 的边界不使用，因为那里可能被弄破或不平整。从表 1 列出来的内容可知，最大使用的载玻片面积为 22 mm × 73 mm。许多商用扫描器仅仅扫描 60 mm 宽的范围。

（七）均匀性

均匀性是对荧光发射与探测连贯性的测量。在这个成像面，光收集的均匀性是共聚焦扫描仪特别关注的。共聚焦的特征决定了聚集的深度，经常是几十微米。这就意味着在扫描运动中距离平面的偏差或在样品中的偏差会导致成像面的非均匀性。增大共聚焦针孔可提高聚集深度，但是会带来不希望的图像瑕疵。共聚焦扫描仪特别是移动样品的扫描仪具有均匀激发整个场的优点。

（八）图像几何学

对扫描仪图像几何学的要求完全依赖于图像数据的后续处理。如果图像定量分析软件使用扫描图像的固定栅格并希望在每个区域内找到样品点，则对图像的大小、X－Y 正交性及像素方位线性关系很大。更复杂的微阵列图像定量分析软件使用动态栅格，无论样品点方位在哪里，都可寻找到每个样品点。显而易见，如果仪器放置阵列样品点很少，更高的扫描精度是没有意义的。

许多用户对于图像大小和线性度在比较满意，而许多第二代扫描器超过了这个精度水平。

随机几何误差也是很重要的，这些误差经常称为抖动，在光栅图像中表现为锯齿状的垂直线。当线到线的非重复扫描超过一个像素大小时，样品点图像边缘变为锯齿状。当每个样品点仅包有几个像素时，在定量分析时会带来一些问题。

在不同波长扫描相同样品得到的多幅图像的重合也是值得注意的。后续处理软件能找到这些样品点，并自动在图像上建造栅格，该软件通常在来自一个通道的图像上创建一个栅格，然后把它应用到不同通道的图像中去，许多阵列扫描仪能把连续几次扫描重合起来。

（九）工作量

用户经常会关心一天可以扫描多少样品，即扫描仪的工作量。使用同一分辨率、大小图像及相同通道数，评价扫描仪工作量才有意义。以 CCD 相机为探测器的扫描仪具有非常高的工作量，但探测弱荧光样品的探测效率比第一代共聚焦扫描仪低几倍。

通过在发射通路上放置分光镜，使用几个探测器、几个信号处理模块及多种波长同时扫描，可以大大增加扫描仪的工作量。但这会相应增加扫描仪的成本和复杂性，尤其是多波长仪器会带来光学系统的干扰。

第一代扫描器对 20 mm × 60 mm 区域单波长扫描，分辨率是 10 μm，需要 5 ~ 15 min，而第二代扫描器更快。

(十)信号源的叠加

信号源的叠加不是对性能的测量,而是一个很重要的概念。在扫描仪显示器上看到的图像或通过后续软件处理过的图像并不是微阵列样品点荧光的简化图像,而是几幅图像的叠加。被扫描目标染料的荧光,是所期望探测到的,而光子统计噪声(这实际是图像的一部分,但看起来象噪声)、其他化学物质或玻璃引起的背景荧光、激光反射光(实际是样品的一个反射图像)、电路噪声是不期望有的。

仅单独考察一幅图像或图像数据,我们不能确定每一幅图像贡献的大小。数据仅仅是对许多像素亮度的收集,而 A/D 转换器并不关心数据来自于哪里。

微阵列之间的比较是一个复杂的过程,不仅仅对单个性能参数进行比较,通常需要对扫描样品进行综合测试。

图 16-14 ScanArray® 5 000 激光共聚焦生物芯片扫描仪

第五节 ScanArray® 型共聚焦微阵列扫描仪

如图 16-14 是美国 GSILumonics 公司最新开发成功的新一代生物芯片扫描系统,它具有独特专利设计的激光共聚焦光学系统,配合高安全性 WINDOWSNT 操作系统,使其成为生物芯片研究体系中不可缺少的利器。该公司已经开发出共聚焦微阵列扫描仪 ScanArray® 的生产线,目前全世界的许多生物技术和遗传基因组分析实验室都在使用这种仪器进行科学研究。

一、扫描仪的集成化

上面各节讲述的知识足够让一个有经验的仪器工程师建造一个实验性的微阵列扫描仪。事实上,许多实验室已经这样做了,许多实验室已经公布自己仪器的性能。但是,微阵列扫描仪最重要的价值在于集成一个完整的、紧凑的、功能强大的、性能可靠的产品。根据用户反馈回来的各种信息和要求,仪器生产商已研制出商用扫描仪能更好的满足研究者要求。

表 16－3　ScanArray® 性能规格

性能指标	ScanArray® 3 000	ScanArray® 5 000
结构	样品载体移动，光源固定，多个激光头，连续扫描	样品载体移动，光源固定，多个激光头，连续扫描
数值孔径(NA)	0.75	0.75
载玻片	所有标准的 1″×3″载玻片	所有标准的 1″×3″载玻片
视野	不小于 22 mm×60 mm	不小于 22 mm×73 mm
图像均匀性	全视场最大 ± 10%	全视场最大 ± 10%
扫描速度	最大每秒 6 线	最大每秒 20 线
分辨率	10 μm 或 50 μm	5、10、20、30 或 50 μm
激光器	2 种(543 nm 和 633 nm)	4 种(488、514、543、593、612 或 633 nm)
发射通道	2 个	10 个以上
主机	奔腾处理器，Window 操作系统	奔腾处理器，Window 操作系统
输出格式	16 位 tiff 格式或 bmp 格式图像	16 位 tiff 格式或 bmp 格式图像
体积	25″×10.5″×11.5″(长×宽×高)	30″×15.5″×13.5″(长×宽×高)
电源输入	电压 95～250 V，频率 47～63 Hz 交流电	电压 95～250 V，频率 47～63 Hz 交流电
重量	35 磅	60～70 磅(取决于激光器数目)

二、ScanArray® 扫描仪结构组成

ScanArray® 扫描仪是一种能移动载玻片、使用稳定光学系统、拥有多种激光器的共聚焦扫描器。它使用简单的光学系统，拥有很大的数值孔径和高收集效率，且成本很低。紧凑的结构能减少实验室占用的空间，拥有这么多优点而不会减少工作量，巨大的努力已研制出可控制样品移动的机械装置，每秒扫描可达 20 线。

HeNe 激光器和氩离子激光器提供不同波长的激发光，一块几何分光镜和特殊设计的发射滤光片能最低限度地降低激光反射光。

用户可以通过主机来控制 ScanArray® 3 000 或通过网络联接控制 ScanArray® 5 000。使用 Windows－95 或 98 或 NT 操作系统，能够快速、方便地操作扫描仪，自动完成多种任务。表 16－3 罗列出其最重要的性能规格。

比较起来，微阵列扫描器是复杂的，许多性能测量是互相关联，选择标准不要局限单一规格，要整体评估。图 16－15 和图 16－16 是使用 ScanArray 仪器获得的一些灰度级微阵列扫描图像。

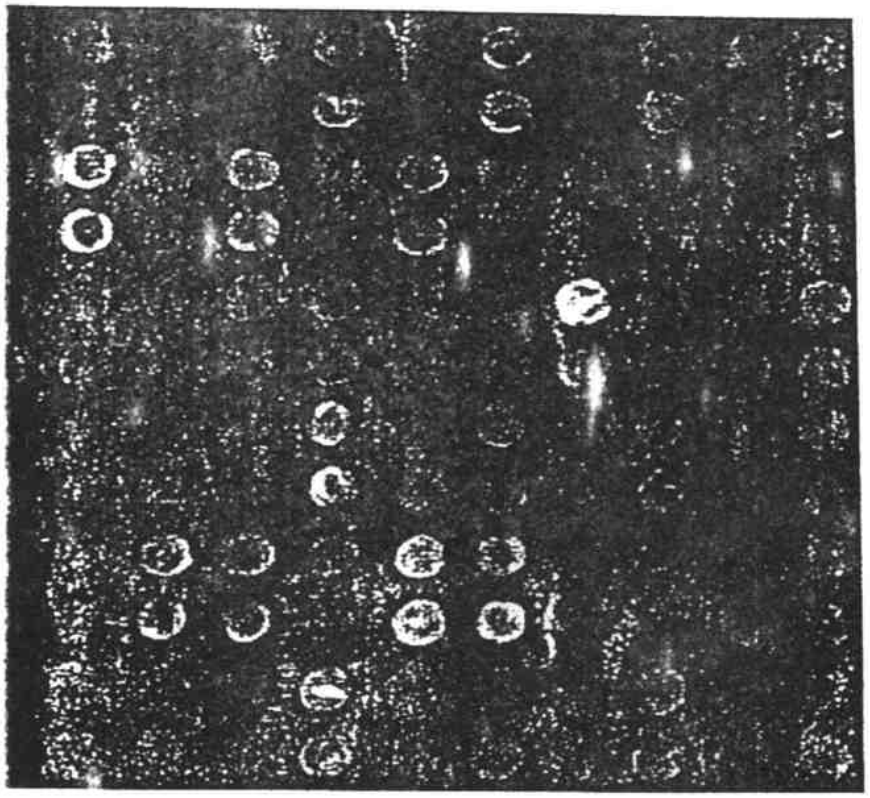


图 16－15　使用 ScanArray® 3 000 扫描基因表达微阵列获取荧光图像

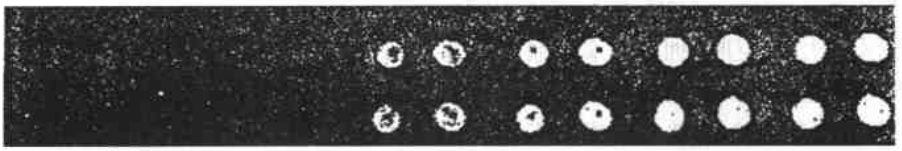


图 16－16　测量系统探测灵敏度(FITC 样品,每组 4 个样品点,从右到左浓度依次降低)

三、ScanArray® 扫描器优点

与其他商用微阵列扫描器相比,新型 ScanArray® 扫描器显示出几种优点:

1. 激光共聚焦扫描光学系统。

2. 全范围高能量激光光源,涵盖最广泛的荧光检测波段。

3. 扫描精度可从 5～50 μm 分级调整。

4. 高灵敏度达 0.1 荧光分子/ μm^2,分辩率高达 5 μm 像素点。

5. 快速全玻片范围扫描时间仅需 5 min。

6. 实时激发及检测光路校正,确保仪器间和仪器内的时间温度最小漂移,全自动信号增益及平衡,荧光信噪比最高。

7. 高度自动化安全的运行及 WINDOWSNT 操作平台,高速局域网数据处理设置,顺应当今

科研潮流，可配置条形码阅读系统进行自动化芯片及基因标识示踪定位，自动进片器省时省力。

8. 极灵活、外形精致节省空间，激光头休眠功能可有效延长仪器寿命。

另外，用户经过少量的训练就能调节仪器获得最大工作量。

（骆清明）

第十七章

生物芯片在医学中的应用

内容提要

基因芯片是在基因组水平上发现并研究基因功能的有力工具。基因芯片技术作为一种高通量、超微量的检测技术，在医学领域中将有独到的优势，是疾病诊断、治疗、预防、和机理研究的有力工具。本章重点介绍了基因芯片在：①遗传病的突变分析和诊断；②单核苷酸多态性(SNPs)检测及其与疾病相关性研究中的应用；③肿瘤基因表达谱的研究、寻找肿瘤相关基因并探讨其功能、肿瘤分型和诊断、筛选抗癌药物和检测肿瘤治疗效果中的应用实例；④肝脏疾病诊疗中的应用；⑤病原微生物快速诊断，特别是耐药性和亚型分析中应用，也为监控新现和再现病原体提供了重要的技术手段；⑥药物作用机制、毒副作用评价、高通量筛选诸方面的应用，特别是指导个性化用药中的应用；⑦法医物证学、法医病理学等方面的应用展望。

第一节　基因芯片技术在基因组功能研究中的应用

一、寻找和发现新基因

为了从众多的文库克隆中找出所需要的基因克隆，传统的方法是采用菌落印迹，十分费时，而且每次只能筛选一个目的基因。而基因芯片技术(DNA chip technique)的效率要高得多。将 cDNA 文库的克隆扩增，以点阵形式排列在载体上，用特定的寡核苷酸探针系统与其杂交，对杂交信号进行分析即可快速地对文库中代表不同基因的 cDNA 克隆予以鉴定和分类。通过相应的公式计算出克隆的相似系数(SIM)，并进行比较，不但可以得到各基因的表达丰度，而且可以发现新的基因。Drmanac 等曾用 cDNA 芯片进行大范围的基因寻找，把来自婴儿脑组织的 73 536 种 cDNA 克隆点在 3 张杂交膜上，分别与数量为 320、230、200 的探针系统杂交。探针中有 1/3 是在编码区出现频率最高的寡核苷酸，1/3 是在非编码区或 ALU 重复序列中出现频率最高的寡核苷酸。杂交后成功地对 19 726 种基因进行了鉴别，还发现文库中约 20 000 个基因属于组织特异性基因。比较 SIM 值发现了约 2 000 个新基因。

1996 年，Schena M 等又应用 cDNA 的微阵列检测了 T 细胞经热休克或佛波酯处理后基因表达的变化，进一步显示了芯片技术在 T 细胞人类基因表达检测、生物学功能研究和基因发现

方面的应用。从外周血淋巴细胞 cDNA 文库中，随机挑取 1 046 个未知序列的克隆，插入片段经扩增后，用高速点样仪连同 10 个对照点在玻片上形成微阵列。分别用两种不同的荧光素标记来自正常培养 T 细胞、热休克处理的 T 细胞(43 ℃温育 4 h)和佛波酯处理的 T 细胞(50 ng 佛波酯处理 4 h)的 cDNA，再与芯片进行杂交。结果在热休克反应中，发现 17 个基因差异表达，其中 11 个受热休克处理诱导表达，6 个受热休克处理表达下调。经测序分析，17 个基因中，5 个表达最高的是 5 种热休克蛋白，另有 3 个为新基因。在佛波酯处理的 T 细胞中，发现 6 个基因表达有差异，其中一个为新基因。

用寡核苷酸芯片也可发现未知基因。Wodickal 等用 4 种共含 260 000 组 25 nt 寡核苷酸的基因芯片覆盖酵母基因组的所有开放阅读框，将酵母细胞分别在营养培养基和基础培养基上生长，杂交结果显示：90% 的基因在两种培养基上都有表达，灵敏度达 0.05～6 个拷贝/细胞，发现有 176 种 mRNA 在两种培养基中有表达差异，包括一些功能未知的基因。

由于基因芯片技术可以同时检测到组织或细胞内多达 10 000 种基因 mRNA 的表达水平，因此与针对单一信号转导途径的研究方法相比可同时采集到信号转导网络中多个节点的基因表达信息。如用异丙肾上腺素(ISO)和血管紧张素 II(AngII)持续注射刺激可引起明显的心肌肥厚，而撤除这种刺激可逆转这一过程。Friddle 等以含 4 000 个 cDNA/ESTs 的微阵列检测经 ISO /AngII 处理的小鼠心室的基因表达，结果发现共有 55 种基因的 mRNA 水平发生了改变。其中 32 种仅出现在心肌肥厚的形成过程中，包括 7 种分泌蛋白，2 种结构蛋白，一些信号分子，如 ECE－2、COMT、PKC 及 9 种新基因；有 8 种基因仅在心肌肥厚的逆转过程中出现；还有 15 种基因如酪氨酸激酶受体 IGF－II、转录因子 Id1、激酶 Jak3 和 MEKK 以及 3 种新基因同时出现在两个过程中。该研究发现了与心肌肥厚及其逆转过程相关的一系列新基因，建立了新的研究靶点。

二、基因表达分析

基因表达分析是基因芯片技术在基因功能研究中最重要的应用方面。人类基因组编码几万个不同的基因，要理解其基因功能需要具备监测大量 mRNA 的实验工具。长期以来对于同时分析多个基因的表达受到技术方法的限制。RT－PCR、Northern－blot 和 RNase 保护实验只能在同一时间内完成对几个基因的检测；差异显示技术难以标准化，重复性差；减数杂交在去除 rRNA 的同时难免造成低丰度 mRNA 的丢失，敏感性降低。1998 年 Antoine 用总 RNA 探针与寡核苷酸芯片杂交对细菌的表达基因进行定量监测时发现其重复性及敏感性均较好。基因芯片技术将大量的功能基因制成表达谱芯片，用不同的荧光染料标记不同组织或细胞的 mRNA，与芯片上的基因杂交、扫描，可得到基因在不同组织或细胞的表达谱图片。再通过计算机分析这些基因在不同组织中表达的差异，可直接快速地同时监测成千上万的基因，这是进行基因功能研究的重要而高效的手段。通过获得的大量信息，可同时对一系列基因表达的个体特异性、组织特异性、发育分化的时空特异性及诱导因素特异性进行综合分析和判断，迅速确立基因与疾病间的关系。

与传统的方法相比，用芯片技术研究基因的表达有如下优点：①微型检测系统，对样品的需要量小；② 同时研究上万个基因的表达变化，揭示基因间的作用关系，效率明显提高；③灵敏度高，可检测低丰度基因的表达。

Lockhart 等对芯片技术定量检测基因表达的敏感性、特异性进行了研究。在阵列的杂交实验中，将 10 种细胞因子的 mRNA 通过体外转录反应掺入荧光素标记，然后随机地将该 RNA 靶

序列裂解为平均 50～100 个碱基大小的片段，再与来源于 B 细胞 cDNA 文库的标记 RNA 混合杂交。标记 RNA 的水平在 1:300 到 1:300 000 之间，40 ℃平行杂交 15～16 h，可重复性地检测出该 10 种细胞因子的 RNA，且杂交强度与 1:300 000 到 1:3 000 之间的 RNA 靶浓度呈线性关系。在 1:3 000 到 1:300 之间的信号则呈现 4 或 5 倍增强。但通过减少杂交时间，线性反应范围可延伸至较高浓度。在小鼠 B 细胞制备的 cDNA 文库中，已知 IL－10 的水平在 1:60 000 到 1:30 000 之间，将 1:300 000 水平的 IL－10 混合到样品中，仍能正确地检测出加入的 IL－10 RNA 量，这提示芯片技术能敏感地反映基因表达中的微小变化。

基因芯片已经可以对某些生物在整体基因组中的水平进行迅速、平行的全基因表达分析。其研究方法如下：①从公共数据库中对预测的开放阅读框(ORFs)序列信息进行查询确认；②识别被确认的 ORFs 序列起始和终止位点两侧的碱基信息；③根据这些碱基序列选择 PCR 引物并自动合成引物序列，对 ORFs 进行扩增；④自基因组 DNA 扩增所有的 ORFs；⑤采用自动化微阵列制片装置将 PCR 扩增产物制成阵列。1999 年 Chamber 等首次利用 DNA 芯片技术对人巨细胞病毒(HCMV)基因组的几乎全部基因的 ORFs 进行了表达监测。用丙氧甲基鸟嘌呤和环己酰亚胺分别封阻病毒 DNA 的复制和蛋白合成，根据基因开放阅读框表达的时间动力学改变进行了分类，描述了立即早期基因、早期基因、早晚期基因、晚期基因在整个 HCMV 基因组中的表达图谱，发现在各个时相之间可能存在调节因子，值得进一步研究。以同样的方式使用高密度 DNA 点阵装置将 2 479 个酵母 ORFs 固定在载体上，可用于与一系列标记的 cDNA 样品杂交，以进行相关的研究。

在人类的基因表达分析研究中，DNA 芯片是系统研究疾病相关基因功能的重要手段，如对肿瘤相关基因的研究。由于细胞间的基因表达差异往往能反映出这些细胞发育正常还是向恶性化方向发展，芯片技术能够提供多个基因差异表达的信息，不但可以确定与肿瘤发生相关的基因，而且可以初步确定可能的作用。其做法是：首先将已知的或可能与肿瘤相关的基因固定于芯片上，这些特定的基因片段来源于 cDNA 文库、ESTs 序列或基因组序列。分别提取正常人群和肿瘤患者的 mRNA 进行反转录，同时荧光标记，与芯片上的基因杂交，可分别得到正常人群的基因杂交图和患者的基因变化图。再用计算机比较它们的基因表达差异。当患者的群体足够大时，就可能得到所有的癌症相关基因。只有通过芯片才可能在短时间内完成如此巨大的工程。通过分析它们在基因组中的位置及与其他生物进行同源比较，可以获得关于这些基因功能的进一步的线索。还可以据此制作点突变检测或肿瘤基因诊断芯片。将 cDNA 差异分析(RDA)与 DNA 芯片联用可迅速筛选出肿瘤细胞中特征基因的差异表达克隆。如为了比较两种具有不同生物学行为的 Ewing 肉瘤组织样品的差异，构建了它们的差减文库。鸟枪法将 RDA 产物克隆到质粒载体上，以克隆的 RDA 产物制作微阵列，然后对来自两种肿瘤组织的 mRNA 的逆转录产物扩增后进行双色荧光标记及杂交反应。发现荧光信号的强度与 RDA 产物及初始 mRNA 的丰度具有很好的相关性。有 50 个单一的差异表达克隆被识别。

心血管系统由多种组织细胞组成，其生理和病理改变的分子机制十分复杂。基因芯片技术可以对其进行多环节、多层次的研究。在用基因芯片对大鼠心肌重塑模型的表达谱的研究中，Stanton 使用了有 7 000 个 cDNA/ESTs 的微阵列。杂交结果显示：经过 2～16 周缺血损伤引起的心肌重塑过程中有 731 个基因的表达发生了改变。多种细胞外基质如胶原蛋白、层粘连蛋白、纤粘连蛋白等的表达明显增高，一些细胞骨架蛋白的表达也有增强，但收缩蛋白的表达减弱，提示心脏的顺应性降低。伴随心力衰竭的出现，一些与脂肪酸代谢相关的酶类表达减弱，而一些与糖代谢相关的酶类表达增强。说明心脏的能量利用模式转为快捷低效的方式。

而与细胞分化、细胞周期相关的基因很少发生表达水平的改变。以上研究结果表明基因芯片技术通过从整体水平提供不同基因的表达信息，可使我们获得特定的生理与病理状态相关的基因谱，包括已知的和未知的。

上述实验说明定量监测大量基因的表达水平在阐述基因功能、发现可能的诊断及治疗靶点等方面是很有价值的。在缺乏任何序列信息的条件下，微阵列可用于基因发现和基因表达检测。既提高了实验效率，又降低了系统误差。目前，大量人类 ESTs 给 cDNA 微阵列提供了丰富的资源。数据库中 400 000 个 ESTs 代表了所有人类基因，成千上万的 ESTs 微阵列将为人类基因表达研究提供强有力的分析工具。这将大大地加速人类基因组的功能研究。

三、DNA 序列测定与序列间比较

1. DNA 序列测定。快速测序是发展 DNA 芯片的初衷。基于凝胶的测序分析实质上都是一维的分离实验，而芯片杂交实验可以很经济地进行大规模的实验。芯片技术中杂交测序（sequencing by hybridization，SBH）技术及邻堆杂交（contiguous stacking hybridization，CSH）技术即是一种新的高效快速测序方法。SBH 技术的原理是将未知序列的靶 DNA 与大量含不同序列的寡核苷酸集合杂交形成完整的双链体，根据碱基配对的原理确定其互补的靶核酸序列。在显微镜下对杂交部位进行鉴定和分析并经计算机分析处理，拼接出正确的靶 DNA 序列。可被分析的目标长度大约等于点阵上寡核苷酸数目的平方根。含 65 536 个八核苷酸的微阵列能确定约 200 个核苷酸长度的序列。SBH 技术的效率随着微阵列中寡核苷酸数量与长度的增加而提高，但同时也提高了微阵列的复杂性，降低了杂交的准确性。CSH 技术弥补了 SBH 技术存在的弊端，又增加了微阵列中寡核苷酸的有效长度，加强了序列准确性，可进行较长的 DNA 测序。计算机模拟证明 8 聚寡核苷酸微阵列与五聚寡核苷酸邻堆杂交，可测定数千个核苷酸长的 DNA 序列。

2. 芯片技术可用于序列比较研究。在遗传分析和基因组分析中，常常需要进行大规模的序列比较，确定序列之间的相似性。利用芯片进行序列比较的原理是：一特定 DNA 在芯片上的杂交特征可与一已知序列的 DNA 片段相比较（杂交特征由短寡核苷酸探针集合与靶 DNA 杂交的分值组成），一个最高得分的特定 DNA 清单（LIST）与已知序列的清单比较用来判断其相似性。这种方式可避免序列重建这个中间步骤，有利于结果识别的专一性。

基因家族各成员间的序列比较分析，对于确定基因的结构、功能、调节和进化等都是很有价值的，而建立在杂交技术之上的序列分析可加快人们对于不同物种的相关序列的研究。Hacia 等利用寡核苷酸基因芯片技术研究了黑猩猩、大猩猩和猩猩等密切相关物种的 DNA 序列。他们发现对 97%以上核苷酸序列相同的高级哺乳动物可获得正确率为 99.91%的碱基顺序。应用二级确认杂交芯片策略，就可以得到它们的定向进化同源基因的全部序列。因此本方法可用于人类基因组多样性计划和进化计划的研究。

杂交测序的不足之处在于：①寡核苷酸集合的组成各不相同，在同一芯片上不可能都处于最佳杂交条件，因此难免出现假阳性和假阴性；②存在碱基错配如 G－T，G－A，难以检测；③SBH 不适于含大量内部简单重复序列单元的 DNA 的序列测定。

四、突变体和多态性的检测

人类基因组是一个高度变异的体系。人类基因组序列图是基于个体的序列，并未反映其变异的一面。为了分析基因的多态性与生物功能和疾病的关系，需要对大量个体的基因型进

行比较、分析。由于大多数性状和遗传性疾病是由多个基因同时决定的，分析起来十分困难，基因芯片技术恰恰解决了这一难题。SBH 技术可大规模地检测和分析 DNA 的变异及多态性。近来又利用结合在玻璃支持物上的等位基因特异性寡核苷酸(ASOs)微阵列建立了简单快速的基因多态性分析方法。又称 DAFT(Direct Allele - Specific fluorescene targeting)。主要步骤是：①将 ASOs 共价固定于玻璃载片上；②采用 PCR 扩增基因组 DNA，其一条引物用荧光素标记，另一条引物用生物素标记；③分离两条互补的 DNA 链，将荧光素标记 DNA 链与微阵列杂交；④通过荧光扫描检测杂交模式，即可测定 PCR 产物存在的多种多态性。用该方法对人的酪氨酸酶基因第 4 个外显子内含有的 5 个单碱基突变进行了分析，结果显示单碱基错配与完全匹配的杂交模式非常易于区别，证明这种方法可快速、定量地获得基因信息。Kozal 等应用基因芯片研究了未曾接触蛋白酶抑制剂的 HIV 患者中 HIV - 1 clade B 蛋白酶的多态性，并用 Sanger 法进行了测序，验证基因芯片所得的结果。总共分析了 114 例样本，发现了在 99 个氨基酸位置上有 47.5%的样本出现多态性，两种方法的一致性达 98%。现在，Affymetrix 公司已制造出商用 HIV 芯片，包括 1 040 个蛋白酶和反转录酶基因，用来研究病毒抗性发展过程中的碱基突变。

高密度探针阵列可检测具有特征性的较长序列的相关多态性与变异。筛选 1 000 个核苷酸序列的变异与多态性需要 4 000 个探针。用 100 μm 合成位点，设计需 1.28 cm^2阵列。将需大约 16 000 个探针，能筛选 4 kb 序列。Incyte 公司的微排序技术结合 Zooseq 数据库中存有的小鼠、大鼠和 Cynomolgus 猴等的基因组序列，能够比较不同生物的基因差异，并对新药的药理学和毒理学进行研究，为临床用药提供线索。

需要指出的是，虽然基因芯片具有快捷、高效的优越性，但目前对其应用也有一定的局限性。①基因芯片不能对待检测基因在多细胞类型组织中的精确定位进行判断，而原位杂交技术可以克服此缺点；②利用基因芯片进行表达分析检测的是转录水平的改变，如果蛋白质调节其功能主要不是依赖其是否表达或表达量高低，而是依赖蛋白质磷酸化 - 去磷酸化等方式，那么在此这种情况下，作为核酸类的生物芯片就没有太大意义，而蛋白类芯片可能会发挥作用；③在研究发育过程中基因表达的空间定位时 FISH 方法较微阵列更有效；④对高度重复序列的检测还应使用传统的电泳方法；⑤亲缘关系较远的生物基因组大片段未知序列杂交测序的准确性不够高。此外，在用芯片进行不同实验时，如何确定一个合适的比较标准，如何准确衡量低拷贝 mRNA 的微小改变，都是现阶段尚未完善的工作。

目前，国外已经有 10 多家从事 DNA 芯片研究的公司，一般的实验室可以根据实验需要选择购买已经商品化的预制芯片。目前已经有不少厂家提供成熟的商品化芯片，如 Affymetrix、Clontech 等公司，产品质量非常可靠。也可以依照用户的实际研究工作的需要，量身定做特殊作用的芯片，设计制备芯片的费用根据芯片上的基因种类、数量、阵列格式、提供信息的价值来衡量。

基因芯片技术要成为实验室研究或临床可以普遍采用的技术仍有一些关键问题亟待解决如：①提高基因芯片的特异性；②简化样品的制备和标记操作；③增加信号检测的灵敏度；④高度集成化样品制备、基因扩增、核酸标记及检测仪器的研制和开发。上述问题是当前和今后一段时期内国内外基因芯片技术研究的焦点。

总之，基因芯片技术虽然还存在这样或那样的问题，但其在基因表达分析、基因诊断、药物筛选及序列分析等诸多领域已呈现出广阔的应用前景。随着芯片技术向自动化、微型化、多样化的发展以及成本不断降低，基因芯片技术的应用将成为我们认识生命本质的捷径。

（曹慧青　李　健）

第二节　生物芯片技术用于单核苷酸多态性的检测

不同种族的人会有不同的肤色和其他外貌特征的差异，还有些人天生具有成为伟大艺术家或是运动员的天赋，而有些人年纪轻轻就得了严重的疾病。存在于人们身上的这些差异实际上都有其遗传基础，DNA 上的差异改变了某些重要蛋白的作用方式。有时这种差异只是 DNA 上的一个单核苷酸的不同，而且许多人同时具有这种差异，我们称之为单核苷酸多态性（single nucleotide polymorphisms，SNPs）。

单核苷酸多态性是指在基因组水平上的特定位点由单个核苷酸的变异所引起的 DNA 序列多态性。具体的说，SNPs 是指基因组内特定核苷酸位置上存在两种或多种不同的碱基，其中最少一种在群体中的频率不应小于 1%。

1998 年国际上就已开展了发现新 SNPs 的研究，目前，不仅在人类染色体上，而且在其他生物的基因组上也已建立了 SNPs 图谱。现在，已鉴定和放入数据库中的 SNPs 已超过 210 万个。我国现在也已启动中华民族基因组单核苷酸多态性（SNPs）系统目录的构建与研究。

研究表明，人类在基因组上的相似率达到 99.99%。SNPs 是人类基因变异的一种最常见形式，占所有已知多态性的 90%以上。SNPs 出现的频率估计是 1/1 000 bp ~ 10/1 000 bp。在人类全部基因组中，可能存在着 300 万 ~ 1 000 万个 SNPs。不同的人群（种）有不同的 SNPs 分布特征，这是他们的种族起源、各种遗传疾病易感性、外貌和生理特征等各方面的差异在基因组水平上的表现。SNPs 对医学研究、临床诊断、临床药物试验、新型药物的开发、物种进化和遗传研究、法医学以及群体遗传学的发展具有重要意义，而且成为新一代遗传标记，可简化现已绘制的庞大的人类基因组图谱，有利于对疾病相关基因定位研究，尤其是对多基因遗传病进行准确基因定位。

理论上讲，由于同一位点可以出现 4 种核苷酸（G、C、A、T），因此 SNPs 既可能是二等位多态性，也可能是 3 或 4 个等位多态性。但通常 SNPs 都是指二等位多态性，因为后两种情况极为罕见。这种基因上碱基的变异大多数为转换，而颠换较少，转换型变异约占 2/3。研究表明 SNPs 出现在 CG 序列上的转换最为常见，原因是 CG 中的 C 即胞嘧啶其中大多数是甲基化的，可自发地脱去氨基而形成胸腺嘧啶，从而形成 C→T 的转换。在基因组 DNA 中，任何碱基均有可能发生变异，SNPs 既有可能在编码区，也可能在非编码区找到。不过，实际上大多数 SNPs 是在非编码区找到，而位于编码区内的 SNPs（coding SNPs，cSNPs）比较少，但鉴定出编码区 SNPs 即 cSNPs 的意义非常重大，因为它往往代表基因型和表现型的改变。

一、SNPs 的检测技术

SNPs 本质上是基因组 DNA 序列中单个碱基的变异，理论上任何用于检测单碱基突变或多态性的技术都可用于 SNPs 的识别或检出。传统的 SNPs 检测方法是采用一些已有的成熟技术，如 DNA 测序、限制性酶切片段长度多态性（RFLP）、单链构象多态性（SSCP）、等位基因特异的寡聚核苷酸杂交（ASO）、寡核苷酸连接分析（OLA）、等位基因特异的 PCR（ARMS）等。不过这些技术大多需要进行凝胶电泳检测，检测过程费时费力，不易自动化，因此未被广泛采用。目前已有多种新方法可用于 SNPs 检测，如 DNA 微阵列的分析法、变性高效液相色谱（DHPLC）、TaqMan 系统、飞行时间质谱（TOF－MS）、侵入测试、动态等位基因特异的杂交、寡聚核苷酸特异的连接等。目前，仍有人不断提出新的检测方法，表 17－2 中对这些方法作了一个小结，在他们的文献报道中许多只是提出了一个大概方法，并没有详细过程。虽然，这些方法都

提到能高处理量地进行 SNPs 基因分型，但是它们成为一个高处理量的 SNPs 基因分型的产品还有待发展，大多数情况下，这些方法只限于学术领域，很难形成实际应用。

（一）DNA 微阵列分析法

DNA 微阵列法的基本原理是：以玻璃、硅、聚丙烯等作为载体基片，在其上进行微阵列分析，让目标 DNA 与大量核苷酸阵列即具有特定碱基序列的探针进行杂交，通过杂交后的荧光检测和计算机分析，便可测出被检测序列的变异。DNA 微阵列的基本过程如下。

1. 寡核苷酸阵列的装配。通常有两种技术用于寡核苷酸阵列的装配。一种是预先合成寡核苷酸后加到载体基片上；另一种是在基片表面原位合成寡核苷酸，两种方法各有其优缺点（表 17－1）。

表 17－1　列阵中寡核苷酸不同合成方法的比较

操作方法	预先合成寡核苷酸后加入	在原位合成寡核苷酸
寡核苷酸合成	一系列	与基片结合
纯化方法	标准方法	不能纯化
适用范围	同一列阵需大量拷贝（用于生产）	大量不同列阵的单拷贝（典型）

在原位合成寡核苷酸，通常有以下几种方法：

（1）利用光平板印刷术（photolithographic approach）合成寡核苷酸，用水银灯将光照在阵列的光平板印刷罩上，从延伸的寡核苷酸链上选择性地移去光不稳定的去保护基团，再循环合成寡核苷酸，直到合成完预先设计的寡核苷酸链。

（2）用机械手臂将碱基合成试剂氨基磷酸脂点至阵列的适当位置上，循环合成预先设计的寡核苷酸链。

（3）用物理方法如掩蔽体来限定前体的位置，将前体通过正交管道只需几步便可合成设定长度的所有相关序列的阵列。前两种方法可以在基片的任意位置上合成寡核苷酸，而后一种方法则不能在任意位置合成不相关的寡核苷酸。阵列的准备是用专用设备在基片表面分割出一系列区带，可系列合成不相关序列的寡核苷酸。如果将一个阵列的寡核苷酸安置在一个平行线上，而将要杂交的样品成行垂直加至寡核苷酸上，这样一次可进行多个反应。在单一的反应中，阵列也能分割出一系列区带。

表 17－2　SNPs 基因分型方法和特点

方法	分辨等位基因	检测	热循环	主要优点	自动化		文献
					分辨等位基因	分析	
微阵列							
1. GeneChips®	杂交	荧光	是	成熟的软件	—	√	Wang et al. [1998]; www.affymetrix.com

续　表

方法	分辨等位基因	检测	热循环	主要优点	自动化		文献
					分辨等位基因	分析	
2. APEX	引物延伸	荧光	是	数据质量	—	√	Shumaker et al. [1996]; Pastinen et al. [1997]
3. OLA	寡核苷酸连接	荧光	是	很高的多路技术	—	√	Khanna et al. [1999]
4. EF 微阵列	杂交	荧光	是	灵活性、严格	—	√	Edman et al. [1997]; Sosnowski et al. [1997]
质谱分析							
5. PNA	杂交	直接	是	原理简单,快速数据积累	—	—	Butler et al. [1996]; Griffin et al. [1997]
6. Masscode™	杂交	Mass tag	是	极高处理量	NIA	NIA	www.qiagen.com
7. Mass Tags	杂交	Mass tag	—	高处理量	—	NIA	Shchepinov et al. [1999]
8. PROBE™	引物延伸	直接	是	准确,快速数据积累	√	√	Little et al. [1997]
9. PinPointTM	引物延伸	直接	是	高质量数据、快速数据积累	—	—	Haff and Smirnov [1997]; Ross et al, [1998]
10. GOOD	引物延伸	直接	是	准确、易处理、不用纯化	√	√	Sauer et al. [2000 a,b]
11. Invader®	核酸内切酶裂解	直接	否	不用 PCR、恒温	—	—	Griffin et al. [1999]
凝胶							
12. MASDA	杂交	荧光	是	容易进行	—	—	Shuber et al. [1997]
13. RFLP	核酸内切酶裂解	荧光	是	容易进行	—	—	Parsons and Heflich [1997]
14. MADGE	PCR	荧光	是	容易进行	—	—	Day and Humphries [1994]
15. SNapShot™	引物延伸	荧光	是	多路技术	—	√	www.appliedbiosystems.com
16. OLA	寡核苷酸连接	荧光	是	等位基因世代的高效因子	—	—	Baron et al. [1996]

续 表

方法	分辨等位基因	检测	热循环	主要优点	自动化		文献
					分辨等位基因	分析	
17. Padlock	连接	荧光	否	不用 PCR	NIA	NIA	Nilsson et al. [1994]; Landegren et al. [1996]
平板记录仪							
18. Invader®	核酸内切酶裂解	荧光 FRET	否	终点，不用 PCR，恒温	(√)	√	DeFrancesco [1998]; Mein et al. [2000]
19. SNP - ITTM	引物延伸	荧光	是	终点，综合系统	√	√	www. orchidbio. com; Syv? nen [1994]; Syv? nen et al. [1993]
20. OLA	寡核苷酸连接	荧光	是	终点，易格式化	—	√	Barany [1991]; Samiotaki et al. [1994]
21. FP - TDI	引物延伸	荧光光偏振	是	实时，同源测试	—	√	Chen et al. [1997, 1999]
22. 5'Nuclease	核酸内切酶裂解	荧光 FRET	是	实时，同源测试	(√)	√	Holland et al. [1991]; Livak et al. [1995]
23. Mol. Beacons	杂交	荧光 FRET	是	实时，同源测试	(√)	√	Tyagi et al. [1996, 1998]
24. Kinetic PCR	PCR	荧光	是	灵活，高处理量	(√)	√	Germer et al. [2000]
流式细胞仪							
25. Coded Spheres	寡核苷酸连接	荧光	是	灵活，高处理量	—	√	Iannone et al. [2000]
26. Coded Spheres	引物延伸	荧光	是	灵活，高处理量	—	√	Cai et al. [2000]; Chen et al. [2000]
其他							
27. Pyro - Sequencing	测序	化学荧光	是	定量	(√)	√	Ronaghi et al. [1996, 1999]
28. Coded Spheres	—	荧光	—	光纤记录仪，高处理量	—	√	www. illumina. com

√：文献中列出；(√)：文献中未列出但理论上可用标准自动化工具；NIA：文献中未直接提到，但暗示如此。

2. SNPs 阵列的装配。最简单的多态性检测阵列类型是在距基片远端的寡核苷酸末端位

置分别合成等位的特异性寡核苷酸，使其末端含可变的碱基。也可以将多态性位点设计在寡核苷酸的中间位置，只是对杂交条件的控制要求更严格，或者在寡核苷酸阵列中使用两种不同颜色的荧光素以区分异源和同源位点，否则易出现假阳性。

3. 靶 DNA 的制备。由于 DNA 样品通常来源于小量的基因组材料，因此需要进行扩增，通常有以下一些方法制备用于杂交的单链 DNA。

(1)先用 PCR 方法将样品 DNA 扩增，然后用核酸外切酶消化不需要的 DNA 链。在合成用于扩增所需链的引物时，其 5′末端的 5 个核苷酸之间存在硫代磷酸化连接，保护该链不被核酸外切酶消化。这样 PCR 扩增后，可用 T7 基因核酸外切酶 - 6 来消化不需要的链。结果不需要进一步纯化即可获得用于杂交的单链 DNA。

(2)在 PCR 反应中，使用一个生物素标记的引物，扩增后的产物经热变性形成两条单链，然后经过一个亲合素柱而使两条单链得以分离，只有生物素标记的链才可结合到柱上。

(3)利用不对称 PCR 方法来制备单链 DNA。在 PCR 反应中，使用的两条引物的浓度比例不一致，一般按 50∶1 ~ 100∶1 的比例加入到反应体系中，在开始的 10 ~ 15 个循环中主要产生双链 DNA，但当低浓度的引物耗尽后，高浓度引物介导的反应将产生大量单链 DNA。

4. 阵列的杂交。寡核苷酸阵列的使用是基于它们能与靶核苷酸以序列特异性方式牢固结合，而使不结合的物质被清洗掉。在最简单的阵列类型中，靶序列通过在阵列中与互补寡核苷酸结合而得以辨认。这样需要非常仔细地控制杂交条件，以使不同的寡核苷酸/靶序列之间有不同的结合能，产生不同的杂交信号。而如果能把杂交信号和用酶使 DNA 序列延伸的过程相结合，通过利用酶来增强阵列的选择性，则杂交条件就不要求如此严格了。根据杂交的先后顺序可有效监测大量 SNPs。在一个 DNA 芯片上可安排一系列短的寡核苷酸覆盖整个 DNA 片段，由于芯片上每一位置的寡核苷酸的精确序列是已知的，杂交的图谱可通过利用荧光标记的 DNA 探针确定下来。这种方法的好处在于可以在一次杂交中扫描一个大的 DNA 片段或一个大的少量的 PCR 产物。

5. 酶催化的延伸反应。有多种可行的杂交/延伸方法适合于单核苷酸多态性分析。在最简单的形式中，合成两种仅末端碱基不同的寡核苷酸，其中一个与待分析的靶序列完全互补。由于对末端碱基的区分能力是很微弱的，两种寡核苷酸和靶序列的杂交以相同的效率进行。而连接酶和聚合酶却能区分这两种双链形式。这两种酶只有在完整双链存在时才能催化序列的延伸。表 17 - 3 总结了在阵列中使用连接酶和聚合酶的相关信息。

表 17 - 3　连接酶和聚合酶在阵列中的使用情况

酶	Polymerase	Ligase
催化反应	利用双脱氧三磷酸核苷酸将核苷酸加到核苷酸链的 3′末端	加入一双脱氧寡聚核苷酸在 3′ - 羟基和 5′磷酸之间形成磷酸二酯键
寡核苷酸与基片的吸附	只有 5′末端吸附到基片上，列阵需要自由的 3′ - 羟基	5′或 3′末端都可吸附到基片上，但 3′端的碱基错配可产生更大的辨别力，所以选择 5′端吸附更好
用于核苷酸或寡核苷酸的标记	荧光素，同位素，生物素	荧光素，同位素，生物素

加入 DNA 聚合酶和带有标记的与靶序列等位位置碱基互补的三磷酸核苷酸，聚合酶将在阵列的寡核苷酸上发挥延伸作用，而和靶序列形成完整的双链。标记物在基片上存在的位置即显示出与靶序列互补的阵列位置(图 17－1A)。在连接方法中，将携有标记的报告寡核苷酸和靶序列加入阵列杂交和连接(图 17－2A)。

在两种方法中，发生变异都是可检测的。寡核苷酸序列在待检测的靶序列的特定碱基上游的一个碱基处终止。在聚合酶测试时，加入某种标记的三磷酸核苷酸，只有当靶序列的特定等位点含有其互补碱基时才能延伸。这一结果还可通过加入其他碱基而得出相反的结果来进一步验证(图 17－1A)。这种测试方法在连接反应中也是可用的，只是用已标记的报告寡核苷酸来代替三磷酸核苷酸，而且这种报告寡核苷酸的末端碱基必须是与待检测的靶序列特定等位点是互补的(图 17－1B)。

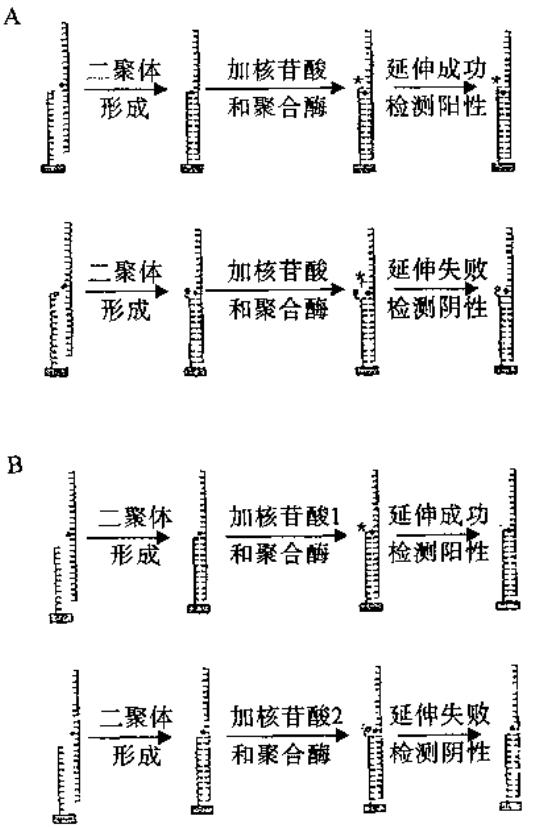


图 17－1　使用聚合酶催化在固定的寡核苷酸上延伸检测 SNPs

A：用一种三磷酸核苷酸在位点特异性寡核苷酸上延伸

B：用位点特异性的三磷酸核苷酸在普通寡核苷酸上延伸

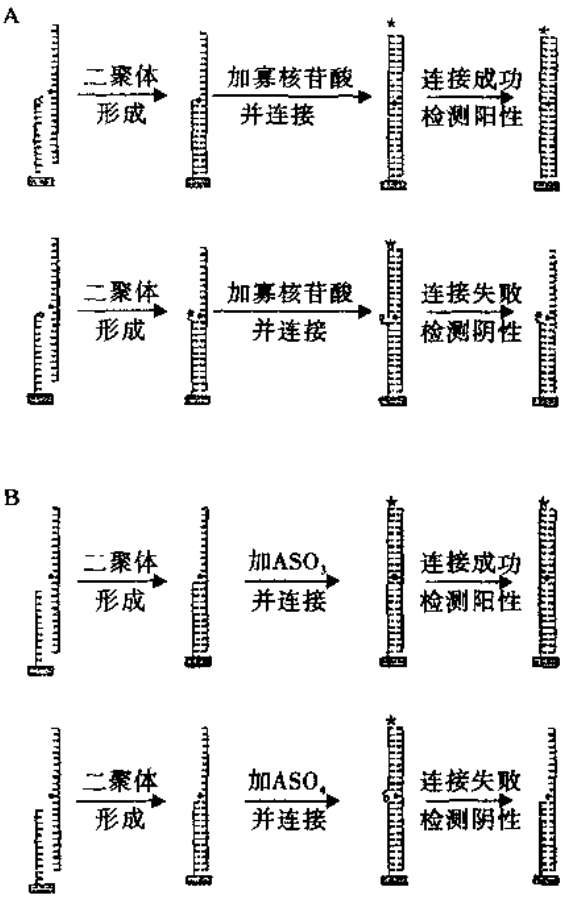


图 17－2　使用连接酶催化在固定的寡核苷酸上延伸检测 SNPs

A:用一普通的报告寡核苷酸与位点特异性寡核苷酸连接

B:用位点特异性的报告寡核苷酸与普通寡核苷酸连接

(二)变性高效液相色谱法(DHPLC)检测基因单核苷酸多态性

运用聚合酶链式反应(PCR)扩增出包含 SNPs 位点的片段,先在 95 ℃变性后再缓慢复性。如果样品中的同源染色体在检测区段是杂合的,即同源染色体的两条 DNA 上一条上存在突变位点,而另一条 DNA 的序列正常,复性后将出现 4 种组分:两种同源双链和两种异源杂合双链,在最佳分离条件下,将可能出现 4 个吸收峰。而如果同源染色体的两条 DNA 链的同一位点均发生突变,即该区段虽然发生变异但是为纯合体,这时可加入正常人的 DNA,这样人为形成异源杂合体的情况,复性后也可得到上述 4 种组分。

然后便可用 DHPLC 来检测 SNPs,基本原理是:将上述复性后的混合样品处理后上样,通过带正电荷的流动相带动使带负电荷的 DNA 分子吸附到固定相上,在 HPLC 运行温度接近 DNA 片段 Tm 值时,含错配位点的杂合双链区域比严格配对的同源配对区更易被洗脱,DNA 分子将根据其与固定相亲和力的强弱被先后洗脱下来,从而出现不同的吸收峰。DHPLC 的优点在于检测时间短、易于自动化操作、而且可用于搜寻未知位点的碱基改变;缺点在于并不能 100% 检测出 SNPs 位点。

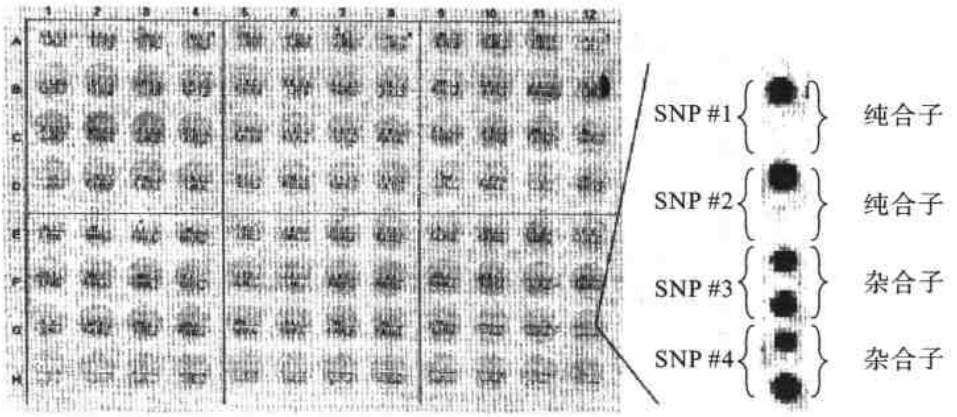


图 17－3　使用 Genometrix's VistaArrayTM 96 孔阵列格式的微阵列基因分型

每一阵列在一平面玻璃片表面有 256 个特异印刷点，针对每一 SNPs 设计野生型和变异型探针并加至阵列中。杂交信号用荧光扫描仪分析。和一个等位基因特异性探针杂交说明是纯合基因型，而和野生型和变异型捕获探针都杂交说明是杂合的基因型。

(三)飞行时间质谱(time of flight mass spectrometry，TOF－MS)

MALDI－MS(matrix－assisted laser desorption mass spectrometry，基质协助激光吸收质谱仪)发展于 20 世纪 80 年代末，已经成为化学、生物学和法庭科学检验分析的一个强有力的工具。

MALDI－MS 技术包括一系列步骤来精确检验 DNA 分子。第一步，制备少量 PCR 扩增产物置于底物表面，通常呈阵列结构；第二步，PCR 产物与适量的含 DNA 样品的有机基质共结晶化以备分析；第三步，产物被精确地检测，基质分子吸收激光发射的短的脉冲导致自发地挥发和电离成基质和 DNA 片段，这种气相离子能够用质谱仪分析。分子量的确定是测量离子的飞行时间(time－of－flight，TOF)，它在质谱仪中与分子的质量成一定比例。这一过程称之为 MALDI－TOF－MS。

利用基质辅助激光解吸附电离－飞行时间质谱技术分析来表征生物分子的特性，将样品溶于一个固体底物中形成晶体，采用脉冲激光轰击，使其离子化，加速离子通过飞行管使其得以分离，所有的离子均可被检测。与激光吸收/电离(laser desorption/ionization，LDI)技术相比，其最大特点是将待分析的分子嵌入小有机复合物组成的结晶基质中，LDI 检测的分子不能超过 2 000 道尔顿，而 MALDI－MS 可以检测分子离子携带的 DNA 片段的几十万道尔顿大小的分子；离子通导率很高，因而具有很高的灵敏度，10～12 mol 的样品即可检测；允许样品中含有少量缓冲液或无机盐；分析速度快，记录一道质谱仅需几个纳秒，十几分钟即可完成一个样品的测试过程。与目前常用的 SDS－PAGE(十二烷基磺酸钠－凝胶色谱)相比，MALDI－TOFMS 不仅测量精度高，而且分析速度快得多。飞行时间质谱(TOF mass spectrometry)以其分析速度快，分析质量范围而在生物学分析中得到了广泛的应用，且无须等位基因梯阶对照，也不需要荧光标记或放射性标记，同时，使用机械系统避免了手工操作检测的不精确性。因此，这一技术提供了一个每天可以检测上千个样品的高通量系统。

MALDI－TOF 质谱方法在分子量可测范围、分辨率和测定准确性方面，具有其他方法不可比拟的优势。然而，这一系统仍然存在一些问题。当样品的 DNA 片段长度或盐离子浓度过高

时，MS 检测灵敏度降低，得不到检测结果。

1997 年 Ross PL 等用肽核苷酸（peptide nucleic acid，PNA）与 PCR 扩增产物进行高特异性杂交后以 MALDI－TOF MS 进行检测，用该方法检测了人线粒体 HVI 区及人 HLA－DQα 基因的 SNPs 多态性。由于 PNA 探针在低盐条件下，与互补的单链 DNA 结合力强，弥补了上述 MALDI－TOF 在此方面的缺点，结合 PNA 探针杂交的高度特异性，使这一方法更加准确和快速。Jiang－Baucom P 也用此方法检测了 HLA－DQα 多态性，实验中以链霉素亲合素包被的磁珠来固定生物素标记的 DNA，使非特异性 PNA 探针能在更强的洗涤条件下被洗涤掉，PNA/DNA/磁珠三者的共轭物再用 MALDI－TOF MS 来检测。

（四）TaqMan 基因分型

TaqMan 等位基因识别测试是利用 *Taq* 聚合酶的 5′核酸酶的活力并检测 PCR 反应中或反应后产生的荧光信号来进行基因分型。TaqMan 基因分型和下面将要介绍的分子信标及侵入测试都是建立在荧光共振能量转移（fluorescence resonance energy transfer，FRET）方法上的技术，FRET 是指发生在两种荧光染料物理上相互靠近时，一种荧光素的发射光谱与另一种荧光素的激发光谱重叠而产生的相互作用。对于 SNPs 基因分型，使用一对 TaqMan 探针和一对 PCR 引物。测试中使用两个在多态性位点不同的 TaqMan 探针，其中一个与野生型等位基因互补，另一个与变异的等位基因互补。将一 5′报告染料和 3′淬光剂染料共价结合到野生型或变异型等位基因的探针上。两个 TaqMan 探针分别用两种不同的报告染料和同一种淬光剂染料标记。如果探针是完整的，由于报告和淬光剂染料的物理作用，荧光被淬灭。在 PCR 退火步骤，TaqMan 探针和靶多态性位点杂交。PCR 延伸期，5′报告，只有完全杂交的探针会被 *Taq* 聚合酶的 5′核酸酶活力所裂解，而错配探针则不会被聚合酶识别。结果导致某一种报告染料的荧光强度升高。通过测量两种不同报告染料在 PCR 反应后的荧光信号强度可进行特殊的基因分型。（图 17－4）这种方法除了可检测 SNPs 外，还可用于检测小的基因删除或插入。

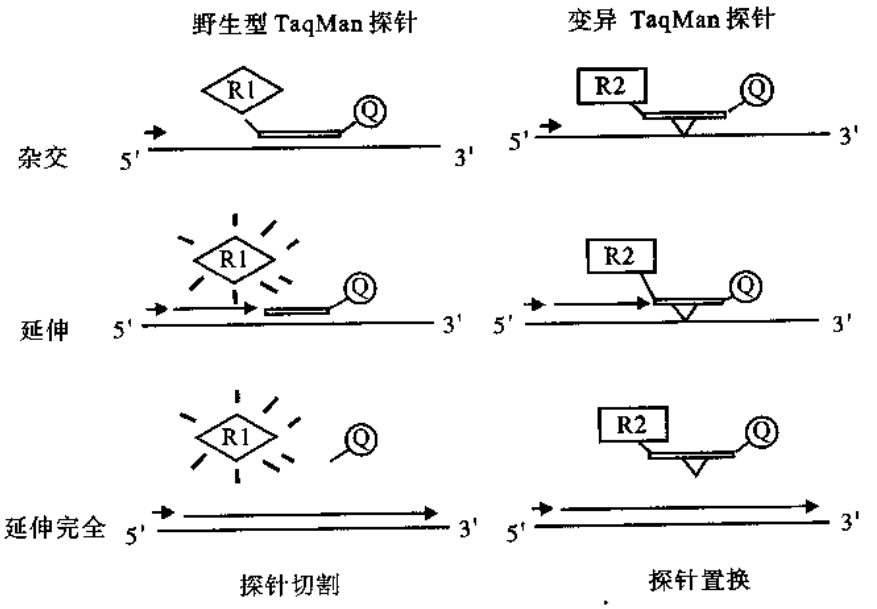


图 17－4　TapMan 基因分型测试

在图 17－4 中，两个与多态性位点结合的 TaqMan 探针用不同的报告染料（R1 和 R2）以及淬光剂染料（Q）标记。在 PCR 延伸反应中，只有完整进行杂交的探针才会被 *Taq* 聚合酶裂解。一错配的探针不会被 *Taq* 聚合酶识别。在基因分型反应的最后，分析 PCR 产物中两种荧光染料 R1、R2 的荧光强度的变化。两种报告染料的荧光比率决定了样品的基因分型。

（五）分子信标（molecular beacons）

分子信标是指两个侧翼序列和靶 DNA 序列互补的的寡核苷酸探针。每一探针的两端有相对的供体——受体染料对。如果不与靶 DNA 杂交，探针自身会形成发夹环，使得报告染料和淬光剂染料靠近，因此没有供体荧光产生。分子信标一旦和正确的靶序列杂交，两种染料分离，荧光强度显著增长。而错配的探针—靶序列杂交，在远低于正常互补杂交的温度时已分离。错配杂交的末端不稳定性增强了分子信标的特异性。对于 SNPs 基因分型，在同一 PCR 反应中，使用两个分别与野生型和变异型等位基因完全配对的分子信标，这两个分子信标分别用不同的荧光素标记。同一 PCR 反应中，利用不同的分子信标可自发检测 3 种可能的等位基因组合。

（六）侵入测试（invader assay）

侵入测试是一种有吸引力的基于 FRET 的基因分型的方法，它在无需 PCR 扩增的情况下即可对 SNP 分型。在每一反应中，使用两个寡核苷酸探针，一个是野生型或变异型信号探针，另一个是上游侵入探针。信号探针包含一完全与靶 DNA 匹配的序列和一荧光标记的侧翼。这些探针以串联方式与基因组 DNA 特殊区域杂交（图 17－5）。当上游侵入探针 3′末端与下游信号探针 5′末端在杂交位点至少有一个碱基对重叠时，该结构会被一种 Cleavase 酶识别和裂解。被裂解的信号探针和模板解离，一个新的信号探针会和模板发生退火作用，随后也会被降解。而如果裂解位点上游有一个单核苷酸的错配，这种结构不会被 Cleavase 酶识别。侵入测试是在恒温条件下进行的，随时间的延长，信号呈线形增长。每一裂解产物于是作为一个侵入寡核苷酸在次级反应中控制着结合有 FRET 标记的探针与模板形成的复合结构的裂解。这种次级寡核苷酸探针（即信号探针）5′末端用供体荧光素标记，荧光则被一内部的受体染料淬灭。当寡核苷酸被裂解，供体、受体染料不再在物理上靠近，淬光作用消除，荧光产生。用一荧光平板记录仪记录每一样品野生型/变异型信号比率便可进行基因分型。

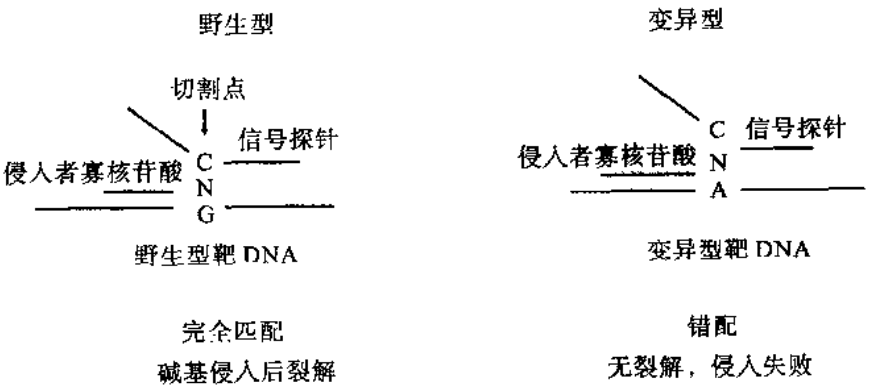


图 17－5　用于突变/多态性检测的侵入测试

侵入测试包含两个序列特异性寡核苷酸——侵入和信号探针的杂交。信号探针含有一完全与靶 DNA 配对的序列和一荧光标记的侧翼。信号探针与侵入探针与靶 DNA 的重叠杂交会产生一个特殊的结构。当侵入探针侵入下游复合结构至少一个寡核苷酸，这一结构将被 Cleavase 酶识别和裂解。被裂解的信号探针和模板解离，于是一个新的信号探针和模板复性，随后也会被裂解。这一过程反复多次，荧光信号呈线性增长。而如果侵入探针和 DNA 模板不能完全匹配，这种侵入作用不会产生。

（七）滚环扩增（rolling circle amplification，RCA）

RCA 是一种非常灵敏的从基因组 DNA 直接检测 SNPs 的技术。对每一 SNPs 设计两个等位基因特异性探针以区分两种等位基因。每一探针的长度为 80～90 个碱基。探针的 5′ 末端被磷酸化并含有 20 个核苷酸序列与靶序列 SNPs 区域 5′端一致，探针的 3′末端含有约 10～20 个核苷酸与靶序列 SNPs 区域 3′端一致。两个探针与靶序列互补区的序列除 3′碱基不同外，其他完全一致，该碱基与两个等位基因的多态性位点是互补的。但这两个探针含有不同类别的主体链序列。一类主体链编码结合两个 RCA 扩增引物的位点，它夹在探针两臂之间。测试的第一步是探针与靶 DNA 结合，靶 DNA 变性后，加入两个探针与其杂交，只有 3′碱基与靶 DNA 多态性核苷酸完全匹配的探针能形成稳定的杂交。完整的杂交使探针的两个末端与靶 DNA 对准，在一耐热连接酶的作用下探针成环。第一个引物与探针主体链的互补区杂交结合，在一链置换 DNA 聚合酶（strand－displacing DNA polymerase）作用下，引物延伸。如果成环的探针形成完整的回转，最终引物会将其自身 5′末端也置换下来。持续的聚合作用和置换作用产生了一个包含起始探针主链区的单链 DNA 连环体。该方法成本低且易于自动化，现在还发展了基于 FRET 方法的 RCA 测试技术，称为 SNIPER™。

二、SNPs 技术的应用

（一）SNPs 在遗传学中的应用

不同的人群（种）有不同的 SNPs 分布特征，这是他们的种族起源、各种遗传疾病易感性、外貌和生理特征等各方面的差异在基因组水平上的表现。

尽管大多数 SNPs 不会造成其所在 DNA 的物理变化，但遗传学家仍急于弄清楚 SNPs 在人的 23 对染色体的分布，其原因主要有两个：①即使 SNPs 本身不会改变蛋白表达和导致疾病，但可能使靠近它的染色体产生有害的突变。由于这种接近效应，人群中可能会存在有害的而尚未被发现的突变，这些人群具有相同的 SNPs，所以 SNPs 可以作为一种标志，这些标志可以帮助发现突变和加速新的治疗药物的发明。②分析 SNPs 在不同人群中的变化，有助于人类遗传学研究人类的演化并找出分散世界各处的不同群体和种族之间的联系。在人类进化过程中，先形成的 SNPs 在人群中通常会有更高的频率。而研究表明不同地域、不同种族人群中特定的 SNPs 并不一定都存在，且所占的比率也各不相同，但大约有 85% 是共通的。例如，通过对高加索人、西班牙人、美国黑人的 SNPs 的遗传多样性研究，结果显示从地理起源区域来看，美国黑人来源于非洲和欧洲，其中 85% 来自非洲地区，15% 来自欧洲地区。

SNPs 的物理定位与微卫星定位相似，但它在基因组中的分布则较微卫星标记广泛，而且 SNPs 是高度稳定的，用其定位的精度比微卫星标记定位要精细，因此 SNPs 已发展成为继限制性酶切片段长度多态性（RFLP）图谱、微卫星标记图谱之后的第三代人类遗传连锁图谱。绘制 SNPs 图谱一般有以下几个步骤：①获取所需的 DNA 序列；②确定序列识别位点（sequence

tagged sites, STSs)；对 STSs 位点进行扫描，筛选 SNPs 位点；③确定并定位 SNPs 位点；SNPs 遗传图谱的建立将进一步完善人类基因组作图。

1998 年，David G. Wang 等开发了对人类染色体组上的单核苷酸多态性进行大规模鉴定、绘制图谱和基因分型的基因芯片技术。在该方法中一个长 L 的靶 DNA 序列可和 8L 大小的变异检测阵列(variant detector array, VDA)中的生物素标记的样本杂交以检测多态性(图 17－6)。对于两条链的每一位置，阵列中都有 4 个相应的 25 bp 的寡核苷酸探针与其对应，4 个探针的区别仅是中间位置的第 13 个核苷酸是 4 种核苷酸中的一种。一个预期序列的纯合子(AA)和完全互补的探针的杂交应该比在中间有一个错配的其他 3 个探针的杂交更强烈。如果存在 SNPs，其纯合子(BB)会针对另一碱基产生强杂交信号而产生不同的杂交图案，而其杂合子(AB)会和两个探针有强杂交。由此，VDA 可检测出某一序列的变异。Wang 等在人类 DNA 2Mb 的序列上选择了 16 725 STSs，其中 1/3 来自随机基因组序列，2/3 来自 3′－ESTs。测试中共使用了 149 个不同的芯片，每个含 150 000 到 300 000 个探针。最后鉴定出 2 748 个候选 SNPs，比率为 1/721bp，其中 STSs 中含有 SNPs 数为 2 299。他们在对该方法错误率的检测后，发现假阳性错误率约为 12%，假阴性错误率为 13%。在此基础上，Wang 等结合基于凝胶的 SNPs 检测法，最终获得了 3 241 个侯选基因，并利用全基因组范围内的 RH(radiation hybrid)绘图法创建了第三代基于 SNPs 的遗传图谱。结合含 STSs 的 SNPs 在该段基因组上的距离为 2.0 厘摩(centimorgan, cM)。

Wang 等随后利用 DNA 芯片技术发展了大规模进行 SNPs 基因分型的方法。他们针对每一待测 SNPs 设计了“基因分型阵列”。每一阵列相应于 SNPs 位置处的两个等位点有两个短的 VDA(图 17－7)。某一等位点的存在可由相应序列的杂交显示的强信号来反映。利用该技术，Wang 等一次完成了对 558 个侯选 SNPs 的基因分型。这 558 个位点各自被扩增、收集、标记并与芯片杂交。为确保有效读出每一位点，他们设计了程式化检测过程，如果位点通过 3 个独立测试的每一个，预期 DNA 序列两条链上的 1 个或 2 个等位点都能成功读出。在所有检测位点中，98%的位点通过了检测，只有 2%由于弱杂交和交叉杂交而失败。

虽然，现在发现的 SNPs 中 cSNPs 所占比例较小，有研究表明 SNPs 在编码区内的变异率仅为其周围序列的 1/5，但由于 cSNPs 的发现往往代表着基因型和表现型的改变，所以其研究意义更重大。cSNP 可分为两种：一种是同义 cSNPs(synonymous cSNPs)，指编码区的 SNPs 并不影响其翻译的蛋白质的氨基酸序列；另一种是非同义 cSNPs(non－synonymous cSNPs)，指编码区的 SNPs 使翻译出的蛋白质的氨基酸序列发生了改变，从而影响了蛋白质的功能。这种改变通常会导致生物表现型的变化。

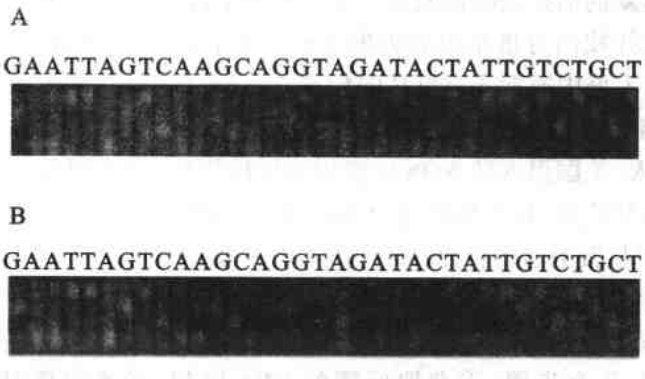


图 17－6　SNPs 检测芯片

图 17－6 中，A 图针对一个 STSs 的 VDA 的一小部分，与预期的靶序列杂交。芯片每一列都是特异的，依次与重迭的 25 bp 寡核苷酸探针结合。同列的 4 行中正中间位置的碱基依次为 A、C、G 或 T，来自预期序列的变异可通过检测每列中最强的信号来确认。B 图是同一 VDA 和在位置 19 处有一 SNPs(A→C)的序列杂交。与上组对照，杂交信号在该位置处更强的是另一碱基，在其附近位置的信号也变弱了(例如，位置 12～18 和 20～26)，这是因为这些位置的探针在对应靶序列 SNPs 处设计的是与等位点 A 互补，而与变异的等位点 C 错配。

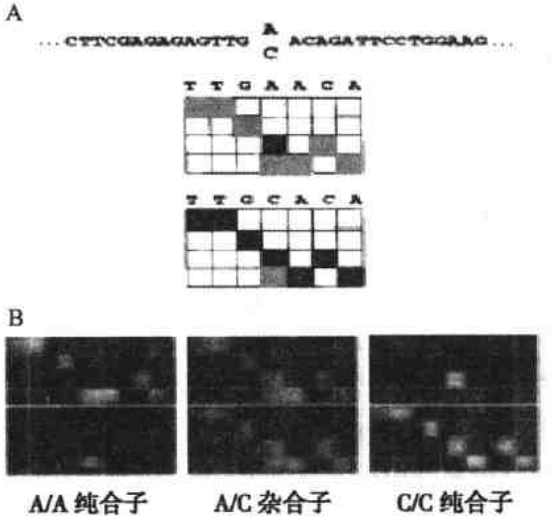


图 17－7　基因分型芯片

图 17－7 中，A 图是一个 SNPs 的基因分型阵列，有二个 VDA 研究 SNPs 及其两旁共 7 个核苷酸，SNPs 位点在中心位置。上下两个阵列各自与含 A 和 C 等位点的序列互补，探针完全和 A 和 C 等位点匹配会显示灰和黑色。针对互补链的基因分型阵列这里未列出。B 图是来自 3 个独立测试的基因分型探针和样品的杂交信号，分别表示 AA、AC、CC 3 种基因型。

(二)SNPs 在医学中的应用

由于 SNPs 在基因组中分布的广泛性，通过对 SNPs 图谱的研究，可以对发病率高的多基因疾病有更深入的了解，对于癌症、高血压、糖尿病、心血管疾病、自身免疫病等复杂的多基因疾病的研究具有重要意义。SNPs 在医学中的应用目前主要集中在以下几个方面：

1.通过多基因分析，寻找和确定疾病相关基因的突变位点。例如，J.Saarela 等采用结合有特殊的 RNA 反转录酶的 SNPs 微阵列技术分析了可能的与多发性硬化症的关联基因区的 SNPs。从基因组 DNA 得到的 PCR 产品含 T3 或 T7RNA 聚合酶转录起始位点，可在体外转录成 RNA，随后用反转录酶转录成 cDNA。将一个 3′端含可能的变异核苷酸的引物吸附到玻片上，如果 3′端和模板序列完全一致，反转录酶将利用 CYP 标记的核苷酸合成 cDNA，用激光扫描仪检测，很容易区分变异和野生型。在实验中，他们设计了两组阵列，第一组 SNPs 微阵列从已存在的数据库中选取位于 5 号和 7 号染色体上的 22 个 SNPs，第二组阵列含来自上述区域的 ESTs 图谱的新的、基因内的 SNPs。同时，对几个位于 17q22～q24 的候选基因如：ICAM－1、PECAM1、髓过氧物酶、Apo 脂蛋白 H 等基因。最终，鉴定出几个和多发性硬化症相关的新的多

态性位点。该方法有助于鉴定预先选定的疾病相关基因。

2.疾病的基因诊断。现在已有如 *P53* 抑癌基因、艾滋病毒、乳腺癌、囊肿性纤维病基因等 SNPs 的芯片或诊断试剂盒。

3.寻找对疾病有调控作用的基因,以期通过新的途径来治疗疾病。

不同的人群或种族对不同的疾病具有易感性,这里有个体与环境的相互作用的因素,更重要的是其基因的本质所决定的。目前有许多方法鉴别基因与疾病的关联,例如,比较疾病和非疾病组织的基因表达图谱;探索人类基因和模式器官在同一疾病中的同源关系;用 in silico 方法从碱基序列揭示相关基因家族的功能。这些方法多数都没有从序列变异形成的遗传差异来研究。现在广泛使用的方法是减数分裂图谱分析。它又可分为两种:连锁和连锁不平衡图谱方法。连锁图谱是用来分析家族内部可能的疾病相关等位基因与等位基因标记之间的关联,它只需要低密度的标记图谱即可。连锁不平衡图谱研究的范围已超出家族范围,因此需要一个高密度的标记图谱。在过去主要是利用微卫星序列进行连锁不平衡分析,而现在则更多地倾向用 SNPs 标志,因为 SNPs 是人类基因组上频率最高的 DNA 序列变异。连锁和关联研究中利用 SNPs 数据库将使人们对复杂疾病的研究有更深理解。高密度 SNPs 图谱将提高研究者在非家族性疾病和药物反应的关联基因中的鉴别能力。

运用 SNPs 技术来寻找疾病相关基因通常分 3 步:

(1)制作一个覆盖全基因组的高密度 SNPs 图谱,在关联分析中,SNPs 的最终价值很大程度上取决于其在人群中的频率分布以及彼此之间的连锁不平衡状况。虽然 SNPs 通常在特定位点上只有 2 个等位基因,携带的信息非常有限,似乎很难满足疾病易感基因精确定位的要求。但这一点可通过加大 SNPs 检测密度来弥补,而完整的 SNPs 图谱完成之后,可以提供远高于此要求的密度。而 SNPs 数量虽多,但因检测速度快,便于自动化控制,所以能较好的用于复杂的多基因遗传病研究。为了能有效地利用连锁不平衡的效力,大量数学分析和计算机模型测试的结果都认为,SNPs 的图谱中的 SNPs 标记位点应至少达到 10 万个。甚至有人认为,1 个具有 50 万个 SNPs 的图谱是不可缺少的。

(2)收集大量正常与患病人群样本,通过对比他们的 SNPs 发生的相对频率,进行关联分析,可以将该疾病相关基因定位与基因组的某一特定区域,再进一步进行序列分析,便可对疾病相关基因精确定位。

(3)利用已知 SNPs 进一步研究其与遗传变化的关系,可以利用固定数目的某种疾病相关基因进行遗传变化研究。如果遗传上 SNPs 变化频率越低,可能预示着它与疾病的产生的关联越大,这样有利于找到与疾病发病初期相关的基因。

目前,寻找疾病易感基因的重点之一是建立数 10 万个 SNPs 标志之间的关联。这些关联分析可以通过大量数据分析来实现。例如,测试患者与健康人的 SNPs 等位点频率或比较一个 SNPs 位点是否比预期的更易或更难遗传给患者的后代。由于复杂的特性通常是由遍布整个基因组的多种效应基因引起,有必要找到分布于不同基因的标志位点,并把这些标记联合在一起分析。传统的多数方法只能一次精确评估一个 SNPs 标志,既确定它在疾病中的边缘效应。这些与疾病有显著相关性的 SNPs 靠近或在易感基因内部,测试每一个 SNPs 与疾病的相关性将可能会造成特异性位点导致假阳性结果(I 型错误)的情况。当同时测试大量 SNPs 并单独分析时,这种 I 型错误很容易被夸大。在多重测试中出现的问题和因此出现的遍布基因组的 I 型错误是目前争论的焦点。这种将每个标记单独分析的方法完全忽略了多基因的复杂特性,没有考虑到易感基因间可能存在的相互作用。虽然有许多学者提出了需要将多种疾病相关基

因联合在一起研究，但能直接用于分析的方法却很少。2001 年，Josephine Hoh 等建立了一种 set - association 方法来评估基因组尤其是不同易感基因的各位置上的 SNPs 标志，并用该方法得到了几个有显著性意义的 SNPs 位点的定位，且控制了全部的 I 型错误。该方法将一个给定的 SNPs、等位基因的关联性、Hardy - Weinberg 不平衡和明显的基因型错误等诸多信息综合，限制了假阴性错误率，提高了测试的可信度。他们将该方法成功运用到病例 - 对照研究中。在对 779 例心脏病人进行血管成形术 6 个月后，342 例显示血管再狭窄，把他们归为病例组，其余人归为对照组。所有个体的基因型在 62 个候选基因上有 89 个 SNPs 标志，用 set - association 方法分析，找到了 9 个基因可能与血管再狭窄易感性有关联，Pmin = 0.040。而如果用单标志分析得到的 Pmin = 0.129，远高于 set - association 方法的 P 值。

在单基因疾病的关联研究中，连锁和连锁不平衡图谱已有许多成功应用的例子，但在多基因相关疾病的研究中，却没有实质性进展。原因是使用减数分裂图谱仍存在一些问题。首先，连锁研究中使用的统计学方法显然无法检测所有基因，除非研究的基因在相关疾病中能产生巨大的影响。而多数同时相关的疾病如肥胖和高血压，都受到许多基因和环境因素的影响，每一因素对疾病易感性和发病只有小的影响。第二，尽管连锁不平衡图谱的检测能力很强，但需要在位于标记位点的等位基因和影响疾病的基因位置之间建立关联。对于基于连锁不平衡图谱的研究而言，了解连锁不平衡有多么强烈以及它在特定群体中从一特定位置延伸多远是至关重要的，这有助于研究者知道一个研究中所需标记的数量和密度。研究表明，大量因素包括从群体起始大小和膨胀率到突变和基因的转变率都影响着连锁不平衡长度。这些因素相互作用造成连锁不平衡长度极大的变异，这种变异不仅存在于不同来源的群体，也存在于不同基因组区域。结果，在不同群体中不同基因组的位置上一系列等位基因中显示的连锁不平衡长度主要依靠历史经验来评估。了解了这些，在现代图谱绘制中评估 SNPs 图谱的利用，应将预期连锁不平衡的特征的传统理论与研究的群体和基因组区域的互补的经验数据结合起来。第三，由于 SNPs 以极高的频率分布于整个基因组，一个 SNPs 图谱对于分析临近某疾病相关基因位置的的一些 SNPs，其密度可能已足够了，但和疾病相关等位基因却显示较弱的连锁不平衡。另外，对于 SNPs 图谱的绘制，发展能评估种群隔离和连锁不平衡现象关系的分析技术和研究方案也是至关重要的。对许多同时相关的复杂疾病而言，发展合适的方法学是很重要的，新的单倍型测试方法、基于距离的绘图方法、将连锁和连锁不平衡结合分析和能适应广泛变化的疾病影响因子研究的随机事件模式，再加上 SNPs 技术的推广使用，也许能为未来研究复杂的多基因疾病开创一个新的局面。

不同的种族或人群对不同的外界环境的适应能力是不同的，在同样的环境下有的人群很容易得病，而在另一种环境下，可能是另外一些群体容易得病，这是由于个体或群体对环境致病因素的易感性造成的。这种易感性的遗传基础是基因组的结构差异或表达差异。研究发现人类对环境致病因素的易感性并不完全是由基因决定的，在环境致病因素作用下的基因表达的差异起着更为重要的作用。将 SNPs 的 DNA 芯片技术用于基因表达研究，在理论上只需很少的样本就可确定环境因素对基因组表达的影响并找出易感基因。

（三）SNPs 与 HLA

人类主要相容性复合体（HLA）定位于人第 6 号染色体短臂 6p21.31，DNA 片段长度约 3 900 kb，占人体整个基因组的 1/3 000。HLA - I 类基因包括经典 I 类基因（HLA - A，HLA - B 和 HLA - C）、非经典 I 类基因（HLA - E，- F，- G）以及假基因等，集中在 HLA 复合体远着丝粒一端；HLA - II 类基因区含有 DR、DQ、DP、DN、DO 和 DM 等基因亚区，位于复合体近着丝点端；

HLA－Ⅲ类基因主要由血清补体成分编码基因 C2、C4A、C4B 和 Bf 组成，位于 HLA 复合体中段，亦称中央区，在此区域内还有一些与免疫功能相关的基因。HLA 基因结构显示高度异质性。迄今，HLA 复合体内共鉴定出 224 个基因座位，其中有产物表达的功能性基因为 128 个，96 个为假基因。

HLA 复合体是迄今已知人体最复杂的基因复合体，有高度的多态性，即在一随机婚配的群体中，染色体同一基因座位有两种以上的等位基因。其多态性与多基因性共同构成了 HLA 的多样性。HLA 复合体紧密连锁，以单倍型的方式由亲代传给子代，因此每个人都有分别来自父母的两个单倍型。HLA 各基因座位上的等位基因均为显性遗传，即每个基因所决定的抗原都在细胞膜上显示。

HLA 系统的另一个重要特点是不同位点上的等位基因之间存在明显的连锁不平衡，即实际观察到的某两个基因出现在同一条单倍型的频率与预期值有差异。不同种族有不同的连锁不平衡。有不同的连锁的不平衡单倍型。

13th IHWC(international histocompatibility workshorp and conference，IHWC)一直致力于绘制 SNPs－MHC 图谱，现已基本完成 HLA－Ⅰ型的 SNPs 图谱绘制，在他们的工作中发现在 HLA 区域的 SNPs 频率比基因组其他区域的要高，HLA－Ⅰ型中，SNPs 频率为 1/100 bp。

针对 HLA 基因区的复杂性，人们发展了一种反相微阵列技术用于 HLA 分型。该方法不是将针对个体 HLA 等位基因的特异寡核苷酸固定，而是将来自不同的个体的 PCR 扩增的 HLA 片段固定在固体表面，用标记的特异寡核苷酸探针与来自不同个体的待测样品同时杂交。这种方法适合于显示大样本的有限数量的重要等位基因，它在遗传连锁研究和检测某 HLA 型与疾病的关联中有较大作用。

现在已知 HLA 与许多自身免疫疾病和移植排斥存在关联。早期的自身免疫疾病与 HLA 等位基因关联的研究是利用血清学方法，近年来采用了 DNA 分型的方法来鉴别，现已验证了 HLA 基因产物在一系列自身免疫疾病包括Ⅰ型糖尿病、风湿性关节炎、乳糜泄、多发性脑脊髓硬化症、溃疡性结肠炎等的易感性中起决定作用。然而，这种关联性方法有时存在较大偏差，许多关联性结果在随后的研究中被认定为假阳性结果。在早期自身免疫疾病与 HLA 等位基因关联性研究表明它是和 HLA－Ⅰ型的位点相关的，而随着研究的进展，显然许多关联是存在于远离Ⅰ型区几百 kb 的Ⅱ型等位基因区。例如，起初认为 HLA B8 和Ⅰ型糖尿病有关，而现在发现它是由 HLA DQ 区的突变引起的。

前面提到了将 SNPs 图谱和连锁不平衡关联分析相结合来检测疾病易感基因的方法，然而，在 HLA 区域，该方法的应用却存在较大困难。由于广泛而强烈的连锁不平衡的存在，疾病相关基因组区域很容易被检测到，然而如何寻找到引起疾病易感性的确切位点和多态性确是一个难题。尤其对大多数 HLA 相关疾病的关联性而言，由于强烈的连锁不平衡使得在多变异导致相同遗传信息的区域，很难将连锁不平衡信息分解到单一 DNA 变异水平，因此如何确定任一个单一多态性在其中扮演的角色是极困难的。这一困难在 HLA 中重复出现且其他一些区域也可见到，例如，在研究Ⅰ型糖尿病时位于 11 号染色体的胰岛素基因的区域以及最近研究克罗恩氏病的位于 5q21 位点的细胞素簇时都存在上述现象。另外，也没有足够大的样本以对一个假设产生令人信服的统计。很少存在自身免疫疾病中被鉴定的遗传位点和 HLA 有较强的关联性。人群分类也是关联性研究造成假阳性的一个因素，如何准确选定对照组样本以便和疾病组形成对比是非常重要的。而由于统计方法的原因，现有的大量 SNPs 也会在对照与病例组人群的不同等位基因变异的多重测试中产生问题。例如，如果对 1 000 个随机 SNPs 进

行关联研究，有人期望在显著性水平 $P<0.05$ 条件下找到 50 个关联，这时某一关联的显著性会由于一个重要候选点在疾病发生过程中的生物学特征或者由于以前数据曾指出某一特殊多态性或基因在疾病中的关键作用而明显增强。这实际是形成一种误导，可能会导致某些数据无法解释。

一些遗传性的单倍型在整个 HLA 区域存在而且有很高的保守性（如 A1、B8、DR3），连锁不平衡使得该区和大量自身免疫疾病以及其他一些非免疫疾病但是由遗传单倍型区域的遗传变异引起的疾病存在关联。在基因组某些地方由于连锁不平衡强烈而单倍型是有限的，因此用少量的单倍型标记即可分析相关的显著性成分，这时，分布于基因组的单倍型图谱可能是大规模研究自身免疫疾病和其他一些失调症的关联性策略的最佳方法。总的来说，广泛分布于基因组的 SNPs 加速了自身免疫疾病易感基因的寻找，然而 SNPs 并不是独立作用的，使用连锁不平衡有利于系统性寻找关联性，但目前方法学上的难题限制了关联性策略的进展。也许在未来几年，将数量巨大的 SNPs 数据库、单倍型特征和遍及基因组的连锁不平衡图谱三者结合起来，并在明确了解各自的统计学限制的基础上，在常见的自身免疫疾病的新的遗传决定因素的寻找中能取得快速进展。

（四）SNPs 与药物的关系

临床用药时，药物对不同病人的疗效有时会产生较大差异，甚至会产生较强的副作用，据全美医学联合会的报道，药物负反应（ADRs）是美国住院及死亡的主要原因之一，导致了每年超过 150 万的住院病例和 10 万例死亡。

造成这种差异的原因在于药物在不同个体体内活化、代谢、清除方面存在的差异，而其本质还是遗传差异，研究表明除了病因、病情、药物相互作用、年龄、营养状况、肝肾功能等原因外，药物转运蛋白、药物代谢酶、药物受体蛋白等药物效应基因的多态性是影响药物代谢与疗效的关键。而 SNPs 能充分地反映个体间的遗传差异。通过研究 SNPs 与个体对药物敏感性或耐受性之间的相关性，可以阐明遗传因素对药物效用的影响，从而对医生有针对性地用药和药物的开发提供指导和依据。1999 年，一些主要的制药公司组成了 SNPs 联盟有限公司，主要目标是发现新的 SNPs 并使之能被公开利用、研究，应用 SNPs 分析来开创一个更安全、更有效的药物治疗的时代。

目前关于单个基因对药物作用的影响方面的研究已经有不少成果。现在已发现一些参与药物代谢的酶的基因或受体基因可以改变药物在体内的代谢和个体对药物的敏感性。例如，药物代谢酶 CYP450 酶系易出现基因多态性，从而造成不同人对药物反应存在显著的个体差异，特别是对安全范围较窄的治疗药物更易产生毒性的差别，如细胞色素 P2D6，其基因多态性会影响到多种药物的代谢，包括 β－受体阻滞剂、抗抑郁药、抗精神病药、镇咳药等。但对于许多复杂疾病来说，单个基因对药物作用的影响甚微，一般不会超过 5%。因此，在基因组水平上进行多基因研究是必要的。SNPs 由于数量众多和易于批量检测，正好为此提供了条件。

目前，正在兴起的药物基因组学致力于研究基因多态性与药物效应的多样性之间的关系，它不仅仅研究单个基因对药物作用的调控，而是将整个基因组作为一个整体来研究它对药物作用的调控，包括单基因的影响、多基因的影响和基因与基因间的相互作用，从而为临床有针对性地合理用药和根据不同基因型群体对药物的反应来改进药物设计提供了理论依据。基因多态性是药物基因组学的基础，主要包括药物代谢酶、药物转运蛋白、药物受体蛋白等基因多态性。药物代谢酶多态性指同一基因位点上具有多个等位基因，其多态性决定了表型多态性和药物代谢酶的活性，造成不同个体间药物代谢反应的差异，是产生药物毒副作用、药效降低

或丧失药效的主要原因之一；药物转运蛋白在药物的吸收、排泄、分布、转运等方面起重要作用，其变异对药物吸收和消除具有重要作用；大多数药物与其特异性受体蛋白相互作用产生效应，药物作用靶点的基因多态性使靶蛋白对特定药物有不同亲和力，导致药物疗效的不同。

在 DNA 或基因结构（基因多态性）上的微小的变异就可能会导致一个蛋白的改变，从而对一种药物产生好的或不良的（药效下降或毒副作用）反应。编码区的 SNPs 位点会导致酶催化活性的改变、蛋白稳定性变化或失活，这些遗传变异还会引起基因错误拼接、基因删除和基因扩增。SNPs 会在药物代谢酶、受体蛋白、转运蛋白的基因中出现，因此 SNPs 提供了一种潜在的工具，作为一种遗传标志来预测蛋白功能的改变而导致的副作用、毒性或对药物的不敏感。例如，和药物代谢酶相关的 SNPs 将会造成个体对药物治疗的 3 种适应结果，和降低药物代谢相关的 SNPs 将导致血清中药物含量上升，预示着副作用或毒性的加强；而和加强药物代谢相关的 SNPs 将导致血清中药物含量下降，从而降低药效。另一个例子是和受体或转运蛋白相关的 SNPs，预示着个体在临床上对药物作用的无反应性，这种结果是因为受体或转运蛋白的作用发生了变化而造成的。

通过对上述基因 SNPs 的研究，将会发现大量与疾病相关的基因位点和药物受体蛋白，发现与药物作用和不良反应相关的基因片段。挽救在临床上由于个别不良反应而被终止研究的药物，研究开发个体化作用的高效药物，为药物开发开拓出广阔的领域。例如诺华公司的治疗精神分裂症药物氯氮平，目前仍是治疗精神病的首选药，但约有 1% 的病人服用后会引起危及生命的粒细胞缺乏症，从面限制了市场，如果能检测出粒细胞缺乏症的药物应答基因，无疑将大大拓宽市场。可以预见，未来的疾病治疗将由传统的药物治疗模型——诊断定向治疗转向基因型定向治疗，即根据个体基因型选择药物和剂量，从根本上改变药物治疗模式，实现个体化医疗。展望未来，个性化药物即根据 SNPs 图谱为不同种族或群体甚至为个人设计的药物有望变成现实。

（五）SNPs 在法医学中的应用

参见相关章节中的叙述。

（朱　科　陈　慧）

第三节　基因芯片技术在遗传性疾病诊断中的应用

遗传性疾病（genetic disease）简称遗传病，通常由于亲代生殖细胞、受精卵或体细胞中某些遗传物质缺失或发生改变，导致子代发病，表现为子代结构异常和（或）功能紊乱，故称为遗传性疾病。目前已发现有 1 万多种，在人群中的发病率约为 5%，我国为 13 亿人口的大国，每年约出生 2 000 万名婴儿，按上述发病率估计，每年大约出生的婴儿中约有 40 万成为遗传性疾病的患者，因此对于广大医务工作者来说，为了预防遗传病患儿的出生，有必要了解遗传性疾病的发病机制及遗传性疾病的诊断方法，基因芯片作为最近出现的基因诊断方法，具有大通量、高容量的特点，因此了解基因芯片及其在遗传性疾病的应用，具有重要的现实意义。

一、遗传性疾病的分类

遗传性疾病通常可分为五大类：①体细胞遗传病；②线粒体基因病；③染色体病；④多基因病；⑤单基因病。

(一)体细胞遗传病(somatic cell genetic disease)

体细胞遗传病的形成以体细胞遗传物质突变为直接原因,最近研究发现部分肿瘤、部分先天畸形以及自身免疫病属体细胞遗传病。如遗传性乳腺癌与体细胞抑癌基因 *Brca* Ⅰ的突变有关,而遗传性胃癌与抑癌基因 *P53* 的突变有关。

(二)染色体病(chromosome disease)

人类正常的体细胞为二倍体,染色体数为 46 条,即 22 对常染色体和 XX 或 XY 性染色体。如果在生殖细胞或受精卵发育早期出现差错,就会发生染色体数目异常或结构畸变,常导致子代智力低下与发育迟缓,称为染色体病,如 21 三体综合征以及 18 三体综合征等。

(三)线粒体基因病(mitochondrial disease)

线粒体为细胞质内一个重要的细胞器,为细胞氧化供能中心,线粒体基因病的突变基因位于线粒体的 DNA 上,其表达与遗传完全不同于核基因,呈典型的细胞质遗传方式。目前已发现 100 多种与线粒体基因突变有关的遗传病,如线粒体脑肌病以及 Leber 遗传性视神经病等。

(四)多基因病(polygenic disease)

多基因病是由多个基因突变所致的遗传病,并涉及到多种遗传因素和环境因素,有家族聚集现象,但无单基因病那样明确的家系传递,目前已发现近 200 余种,一些危害严重的常见病,如高血压、冠心病以及哮喘等均是多基因病。

(五)单基因病(monogenic disease)

单基因病是由于单个基因突变所致的遗传病,通常呈现特征性的家系传递模式,在一对同源染色体上,可能其中一条携带突变基因,也可能两条染色体对应基因座位都是突变基因。依照突变基因所在染色体以及基因显隐性的不同常分为常染色体显性遗传病、常染色体隐性遗传病、X 连锁显性遗传病、X 连锁隐性遗传病以及 Y 连锁遗传病等,一些危害严重的遗传病,如地中海贫血、血友病以及血红蛋白病均为单基因病。

二、遗传性疾病与基因突变

(一)遗传性疾病的基因突变类型

遗传性疾病是由于遗传物质发生突变所引起的疾病,也就是说,遗传性疾病发生的物质基础是遗传物质的突变,包括基因突变与染色体畸变,除了染色体畸变导致染色体病外,其他四类遗传性疾病均与基因突变有关,可见基因突变在遗传性疾病的发病中起重要作用。基因突变系 DNA 分子结构的化学变化,这种变化可以人工诱发,称人工诱变,电离辐射、化学刺激以及病毒也可引起;另一类是自然突变,是由于 DNA 复制、交换、修复损伤等过程中碱基配对"误差"而产生,常分为以下几种类型。

1.点突变(point mutation)。点突变又称为单个碱基的替代,指 DNA 大分子上一个碱基的变异,可分为转换(transition)与颠换(transversion)两种类型,嘌呤取代嘌呤或嘧啶取代嘧啶,称为转换,包括 A－G、G－A、T－C、C－T 4 种形式;嘌呤取代嘧啶或嘧啶取代嘌呤,称为颠换,有 8 种颠换形式。颠换最典型的例子就是镰状细胞贫血,患者血红蛋白分子中的β链第 6 位氨基酸,原本是谷氨酸,由 GAG(G)编码,当 GAG 突变成 GTG,谷氨酸就被缬氨酸所取代,形成镰状细胞型血红蛋白(HbS),发生镰状细胞贫血。

2.插入(insertion)或缺失(deletion)。插入(或缺失)分为两种类型,一是整个密码子插入

(或缺失),此时在插入(或缺失)的密码子上下碱基序列都不改变,其结果是合成的多肽链增加或减少一个氨基酸,如 Hb leiden 病,是由于血红蛋白链的第 6 或 7 位的谷氨酸缺失所致。而另一类指在编码序列中插入或丢失一个或数个碱基,造成编码区阅读框架的移位,造成移码突变,肽链合成因而可提前或移后终止,中国人很常见的 β 地中海贫血突变类型 β codon 41 - 42 (- NJ4bp)就是此类突变。

3.不等交换(unequal crossover)。在减数分裂时,部分同源染色体进行配对,交换遗传物质,如果配对不准确,则可发生不等交换,造成基因或基因片段的缺失、重复等。如两条 16 号染色体的 α 珠蛋白基因的同源序列发生不等交换,可导致 α 珠蛋白基因的缺失或重复,前者发生 α 地中海贫血,后者则可产生 α 珠蛋白基因的三体型。

4.基因倒位(gene inversion)。DNA 分子内一段核苷酸方向反置,如从 5′ - 3′整段颠倒为 3′ - 5′或者 DNA 分子内重组,导致一大段核苷酸在 DNA 分子内移位。重型血友病 A 患者中,Ⅷ因子基因内含子 22 内的一大段核苷酸与因子 Ⅷ 基因上游约 500kb 处的另两段核苷酸之一发生同源重组,从而在染色体 Xq 的末端引入一个倒位,把 Ⅷ 因子编码基因分成相距约 500 kb、方向相反的两部分,导致重型血友病的发生。

5.基因融合(gene fusion)。基因融合系指两个不同基因的核苷酸序列相连接,成为一个新的基因,称为融合基因。其主要原因一是同源染色体之间的不等交换所致,如 β 珠蛋白基因的 3′端与类 β 珠蛋白基因簇中 δ 基因的 5′端融合形成 δ - β 基因,导致 δ - β 或 β 地中海贫血;另一是由于染色体之间易位产生的,如慢性粒细胞性白细胞的 *BCR - ABL* 融合基因,是由第 9 号染色体长臂上的癌基因 *ABL* 易位到 22 号染色体的长臂,与断裂簇区域基因 *BCR* 相连接,成为 *BCR - ABL* 融合基因,该融合基因翻译成一个异常的 *BCR - ABL* 融合蛋白,导致慢性粒细胞性白细胞。

6.动态突变(dynamic mutation)。动态突变又称不稳定核苷酸重复序列突变,指 DNA 分子中核苷酸重复序列拷贝数发生不同程度的增加所致,表现为突变基因的核苷酸重复序列的异常扩展及在代与代之间传递中的不稳定性。如核苷酸重复序列的扩展超过一定的长度就会发生疾病。近年来,陆续发现 20 余种遗传病与动态突变有关,如脆性 X 综合征与 CGG 重复序列异常扩展有关、而亨廷顿舞蹈病则与 CAG 重复序列异常扩展有关。

(二)基因突变的遗传效应

1.遗传密码的改变。由于基因突变,导致核苷酸的序列发生改变,有可能导致遗传密码的改变,根据遗传密码是否改变,可将基因突变分为 4 类:

(1)中性突变(neutral mutation):中性突变是指碱基替换后虽然一个遗传密码子变成另一个密码子,但由于遗传密码存在简并性,其所编码的氨基酸并不改变;或编码的氨基酸虽然改变,但蛋白产物的活性并不改变(如某些同功酶的氨基酸组成不同,但功能相同),如 CUU、CUG 以及 CUC 都是编码亮氨酸的密码子。

(2)错义突变(missense mutation):错义突变是指碱基替换后,一个密码子变成另一个密码子,编码的氨基酸改变,导致蛋白产物的活性改变。镰状细胞贫血是由于 β 珠蛋白基因 codon 6 GAG 突变成 GTG,谷氨酸突变为缬氨酸,即为典型的错义突变。

(3)无义突变(nonsense mutation):无义突变是指碱基替换后,一个原本编码氨基酸的密码子变成一个终止密码子,导致多肽合成提前终止,产生无活性的多肽片段。如 Hb Mckees Rorkf 变异型,就是由于 β 珠蛋白基因 codon 145UAU(编码酪氨酸)突变为终止密码子 UAA 所致。

(4)移码突变(frame shift mutation):在编码序列中插入或丢失一个或数个碱基,造成编码区

阅读框架的移位，肽链合成因而可提前或移后终止，如 Hbanston 是由于 β 珠蛋白基因 codon 145UAU 第一个 U 缺失，即为移码突变。

2.遗传表型的改变。

(1)分子病：由于基因突变导致其编码的蛋白质分子的结构或功能发生改变，或者蛋白质的缺失，从而引起一系列的病理变化，称为分子病，血红蛋白病以及地中海贫血即为典型的分子病。

(2)先天性代谢病：由于基因突变导致其编码的酶分子的结构或功能发生改变，或者酶含量的降低，从而引起遗传性酶缺陷与代谢紊乱，称为先天性代谢病，如白化病与葡萄糖－6－磷酸脱氢酶(G6PD)缺乏症等。

(3)肿瘤：体细胞内的基因突变是导致肿瘤发生的重要原因，遗传性视网膜母细胞瘤就是由 Rb 基因突变所致。

三、基因突变检测技术

目前实验室常用于基因突变检测的技术如下，其核心技术为核酸分子杂交与聚合酶链式反应。

(一)核酸分子杂交

核酸分子杂交是一种最经典的基因突变检测技术，根据已知的基因突变位点的核苷酸序列，设计相应的寡核苷酸探针，探针与待测的 DNA 进行分子杂交，根据信号的有无与强弱判断待检测 DNA 是否发生突变，包括 Southern 印迹、点杂交以及原位杂交等几种类型。

(二)聚合酶链式反应(polymerase chain reaction，PCR)及其衍生技术

用于扩增位于两段已知序列之间的 DNA 区段。在由 *Taq* DNA 聚合酶催化的一系列合成反应中，使用两段寡核苷酸作为反应的引物。引物分别与模板 DNA 两条链上的一段序列互补，而这两段模板序列又分别位于待扩增 DNA 区段的两侧。反应时，首先在摩尔数大大过量的两段寡核苷酸及 4 种 dNTP 存在下将模板 DNA 进行加热变性。随之将反应混合物冷却至某一温度，使引物与它的靶序列发生退火。此后退火引物在 *Taq* DNA 聚合酶作用下得以延伸。如此反复进行变性、退火和延伸称为 PCR 的一个循环。由于一轮扩增的产物又充当下一轮扩增的模板，所以在这一周而复始的过程中，每完成一个循环，理论上使目的 DNA 增加 1 倍。PCR 操作简单、特异性强、灵敏度高，常借助于聚丙烯酰胺凝胶电泳与毛细管电泳进行结果判断，广泛应用于多种遗传病、肿瘤、感染性疾病等的基因诊断。

PCR 扩增产物又可以作进一步的分析，下面将分别介绍。

1.单链构象多态性(single strand conformation polymorphism，SSCP)。基因发生突变时，即使是一个碱基的变化，也会使单链 DNA 的构象发生变化，进而导致单链 DNA 在非变性凝胶中电泳时迁移率发生改变，比较标本和野生型对照基因在非变性聚丙烯酰胺凝胶电泳时迁移的距离或条带的差异，就可判断是否发生突变，但是不能明确突变的类型。

2.切割片段长度多态性(cleavage fragment length polymorphism，CFLP)。CFLP 是在 SSCP 基础上发展起来的突变检测技术，单链 DNA 在中性条件下可形成发夹状二级结构，cleavase I 外切酶能识别发夹结构并可在其附近进行切割。发夹结构的数目是由其一级结构决定的，因此当基因发生突变时，其单链 DNA 形成的发夹结构的数目亦随之改变，通过聚丙烯酰胺凝胶电泳就可检测出来。

3.PCR 与其他技术相结合。

(1)PCR/ASO(allele specific oligonucleotide,等位基因特异性寡核苷酸)探针法。首先采用 PCR 技术扩增靶 DNA,然后将扩增的 DNA 与相应的等位基因特异性寡核苷酸探针进行杂交,严格控制杂交条件与洗脱条件,使只有与探针完全互补的靶 DNA 才显示杂交信号,用于判断是否存在基因突变以及突变是纯合子还是杂合子。

(2)基因芯片(genechip)技术。针对不同的基因突变或基因多态性,设计出不同的寡核苷酸探针,并固定在固相载体上,即基因芯片,同时采用 PCR 技术对含突变或多态性位点的基因片段进行扩增,将扩增片段与基因芯片进行杂交。从某种意义上,基因芯片技术是 PCR/ASO 技术的集合,是大通量、高容量的 PCR/ASO 技术。

(3)荧光共振能量转移(fluorescence resonance energy transfer,FRET)结合探针熔解曲线分析技术。荧光共振能量技术采用了两个不同的探针,分别称为锚探针与报告探针,锚探针的 5′端标记荧光素分子,而报告探针 3′端标记另一荧光素分子(该探针设计与突变型匹配,覆盖靶序列的某一突变或多态性位点),由于两条探针能与模板链相邻的序列杂交,借助于荧光共振能量传递,锚探针所发出的荧光被报告探针所吸收,报告探针发出另一波长的荧光,也就是只有两条探针均与模板链杂交时,报告基团才发出荧光,因此在扩增完成后,逐渐提高温度,对杂交产物的稳定性监测,得到熔解曲线,当反应体系的温度达到解链的温度 Tm 时,探针与靶序列的解链,此时反应的荧光强度变化最大,由于突变型与报告探针匹配,而野生型的靶序列与报告探针有碱基误配,因此两者 Tm 有差别,因此通过 Tm 的变化就可反应待测序列是否有突变以及突变的类型。

PCR - FRET 技术相结合的分析方法就是目前广泛应用的各种定量 PCR 技术的理论与方法学基础。

(4)限制性内切酶酶谱分析。基因突变可导致基因上某一限制性内切酶酶切位点的改变、消失或增加,如用此酶消化待检测 DNA 和野生型对照 DNA,通过比较片段的大小以及多少的差别,判断 DNA 是否发生突变。

(5)限制性片段长度多态性(restriction fragment length polymorphism,RFLP)分析。个体间基因的核苷酸序列存在着差异,为 DNA 多态性。DNA 多态性可影响限制性内切酶的切割位点,而某些致病基因可能与特定的多态性 DNA 限制性片段紧密连锁,以此多态性片段为"遗传标记",可判断待检 DNA 的突变情况。

(三)凝胶电泳技术

1.变性梯度凝胶电泳(denaturing gradient gel electrophoresis,DGGE)。基因发生突变时,其稳定性发生变化,熔解温度(Tm)随之改变,进行变性梯度凝胶电泳(甲酰氨的浓度呈梯度递增),突变 DNA 与野生型 DNA 出现不同步的部分解链,进而导致不同的电泳迁移率。

2.毛细管凝胶电泳法(capillary gel electrophoresis,CGE)。毛细管凝胶电泳法是以高压电场为动力,以毛细管为通道的凝胶电泳技术,与普通的电泳技术比较,具有更高的精密度与分辨率,分离 180 ~ 200 bp 的序列效果更佳。

(四)DNA 序列分析

理论上,DNA 序列分析是进行突变检测的最直观和准确的方法,目前 DNA 测序已成为检测突变并最终确定突变类型和突变位点的一种有效方法。

表 17-4　主要基因突变检测方法的比较

方　法	原　理	关键问题
核酸分子杂交	错配时杂交信号减少或消失	选择合适探针，优选杂交条件
RFLP	基因变异导致酶切位点	选择合适限制性核酸内切酶
DNA 序列分析	全序测定，比较差异	工作量大
PCR	变异引起扩增产物不同	选择适当引物
SSCP	变异引起核酸单链构象改变	选择合适 DNA 片段
荧光共振能量转移结合探针熔解曲线分析技术	突变导致探针与模板稳定性发生变化	选择适当的荧光素标记的探针
切割片段长度多态性	突变引起单链 DNA 发夹状二级结构数目改变	选择合适片段
基因芯片	错配时荧光强度减少或消失	设计适当的探针阵列
毛细管凝胶电泳	多元杂交后分析杂交差异	部分或全长基因的扩增

四、遗传性疾病基因芯片

(一)遗传性疾病基因芯片阵列的设计

基因芯片最关键的技术是探针阵列的设计，不同的靶序列常采用不同的探针阵列设计策略，当进行遗传性疾病基因突变或基因多态性检测时，常采用以下几种方案：

1. 叠瓦阵列(tiled array)，又称砌砖式阵列，是一种最常用的检测遗传性疾病基因突变或基因多态性的阵列，为了检测某一基因突变或基因多态性的位点，一般设计 4 条探针，该阵列探针一般长约 15～25 bp，除了探针最中间的一个碱基外，4 条探针其他碱基序列完全相同，而 4 条探针中间的碱基则分别 A、T、G 和 C，当此 4 条探针与靶序列杂交时，只有碱基完全配对发出的荧光信号最强，根据荧光强弱就可以确定该处的碱基是否发生突变，同样对于一个长度为 L 的待测序列，只要按此方法设计 4L 个探针阵列，就可检测所有的突变位点，例如为了检测序列 AAGGCCATGCXTTACATACAC 中 X 位点是否发生突变，可设计如下探针阵列：

TTCCGGTACGTAATGTATGTG
TTCCGGTACGAAATGTATGTG
TTCCGGTACGGAATGTATGTG
TTCCGGTACGCAATGTATGTG

2. 根据基因突变的类型，如缺失、插入或点突变，直接设计与突变序列互补的探针(突变探针)，同时设计与正常序列互补的探针(参考探针)，组成一个阵列，然后将待测序列与阵列进行杂交，可根据荧光信号强度的强弱确定是否发生突变以及基因分型，如正常序列为 TTCCGGTACGTAATGTATGTG，而突变序列则为 TTCCGGTACGGAATGTATGTG，则相应的探针分别为 AAGGCCATGC[A]TTACATACAC 和 AAGGCCATGC[C]TTACATACAC，我国深圳赛尔公司推出的 Thalachip 地中海贫血基因诊断芯片即为此种类型。

3. 杂交测序(sequencing by hybridization，SBH)用探针阵列。该阵列探针较短，约 7～8 个碱基，包括所有 4 种碱基出现的可能排列组合，以 8 个碱基的阵列为例，有 $4^8 = 65\ 536$ 中组合，以

一定规律相互重叠排列，通过大量阵列反应的结果，就可以判定待测靶序列，当然也可用于了解遗传性疾病的基因突变或多态性位点。而且该阵列还可用于发现新的基因突变或多态性位点，如为了检测 ATCCTAGCTGAA，共有 5 条探针可与待测序列互补，分别是：TAGGATCG、AGGATCGA、GGATCGAC、GATCGACT、ATCGACTT，如果待测序列突变为 ATCCTACCTGAA，则完全互补的探针分别为 TAGGATGG、AGGATGGA、GGATGGAC、GATGGACT、ATGGACTT。

4. 邻堆杂交(contiguous stacking hybridization，CSH)阵列。较短的探针(≤5 bp)不能与靶序列形成稳定的双链，但是将一条较长的探针与一条较短的探针同时与靶序列杂交，且两条探针相邻，由于邻近碱基的堆积作用，则都会形成稳定双链，如相邻的一条 5 bp 的探针和一条 10 bp 的探针同时与靶序列杂交，会形成稳定的 15 bp 的双链。此时 10 bp 的探针与非突变区杂交，起增加双链稳定性的作用，而 5 bp 的探针根据突变的特征设计，与突变区域杂交，因此根据 5 bp 的探针类型以及对应的荧光强度的强弱即可判断突变类型。如为了检测序列 AGCGTGAGCTTAGAT[G]GCTTGCTAT 中[G]是突变为[T]还是缺失，可考虑设计 10 bp 的探针 CACTCGAATC，而 5 bp 的探针则可分别设计为 TAACG 和 TACGA。

5. 通用探针(universal probe)阵列。该探针长约 20 bp，序列一致，也就是说相同的探针可用于不同的靶基因检测，如通用探针阵列可用于不同的遗传性疾病的突变检测，其特异性主要通过 PCR 等手段进行控制，同时在 PCR 等过程中引入探针的互补序列，以便扩增产物能与固相载体连接，如采用通用探针阵列与 PCR/LDR(连接酶链式反应)检测基因突变，首先采用 PCR 技术进行扩增，然后进行 LDR，LDR 的一条引物含有通用探针的互补序列，另一条引物则连接一荧光报告基团，两条引物只有与待测序列完全匹配时，连接酶才能将两条引物连接起来，一旦靶序列发生突变，连接酶不能将 2 条引物连接起来，荧光报告基团不能与固相载体偶联，荧光强度明显下降，因此可用于基因突变的分析，而且可明确基因型，图 17－8 为其原理示意图：

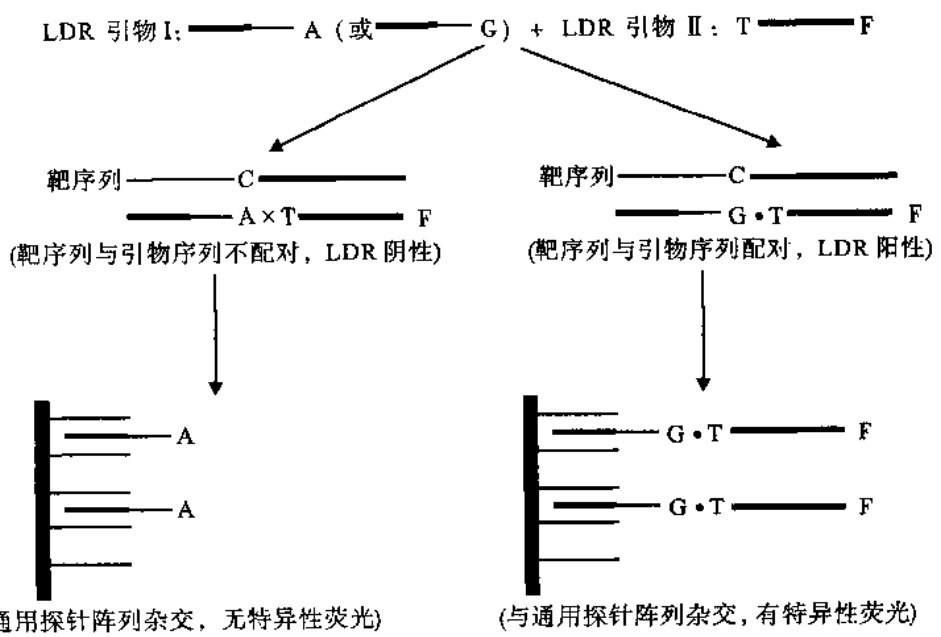


图 17－8 通用探针阵列杂交原理示意图

(二)目前检测遗传性疾病的主要基因芯片种类

1.地中海贫血基因突变检测芯片。地中海贫血(thalassemia)，又称海洋性贫血，是由于珠蛋白基因的缺失或缺陷，导致一种或几种珠蛋白肽链合成受到抑制，进而引起的溶血性贫血，因本病最早发现于地中海地区，因而称为地中海贫血，在我国多见于南方各地，如广东、广西以及海南等省。根据珠蛋白肽链缺失的差异可将地中海贫血分为α地中海贫血与β地中海贫血两大类。

(1)α地中海贫血(α－thalassemia)是指珠蛋白α链部分或完全缺失的地中海贫血，α珠蛋白的基因位于16号染色体短臂上，总长约29 kb，共包括7个连锁的α类基因或假基因，α地中海贫血的基因突变主要为基因缺失，目前至少已报道了20多种缺失型，对中国人来说，α类珠蛋白的基因缺失型主要为右侧缺失(－$\alpha^{3.7}$)、左侧缺失(－$\alpha^{4.2}$)和东南亚缺失型三种。除基因缺失外，尚有部分α地中海贫血由点突变所致，目前已发现的点突变有30多种，中国人主要有两种，一为α2珠蛋白基因中125位密码子CTG突变成CCG，使亮氨酸突变成脯氨酸，另一种则是α2珠蛋白基因中142位终止密码子TAA突变成CAA，使肽链合成直至173位的另一终止密码子出现才结束。

(2)β地中海贫血(β－thalassemia)是由于β珠蛋白基因簇的突变，使β类珠蛋白的合成受到抑制的地中海贫血。β地中海贫血呈常染色体隐性遗传，与α地中海贫血主要由基因缺失所致不同，β地中海贫血全世界已发现120多种基因突变类型，除少数几种为基因缺失外，绝大部分是点突变的结果。如IVS2的3个突变，IVS2－645(C－T)、IVS2－705(T－G)和IVS2－745(C－G)均为点突变，中国人有15种。最常见的突变类型有六种，分别是Codon 41－42(－4bp)、IVS2－645(C—T)、Codon 17(A－T)、Codon 71－72(＋A)，IVS1－5(G－C)和TATA box －28(A－G)。

地中海贫血是发病率最高的遗传性疾病之一，1989年，Saiki RK等根据HLA－DQA等位基因和珠蛋白基因突变的特点，分别设计6条与9条探针，固定子尼龙膜上，开创了基因芯片用于地中海贫血诊断的雏形。随后Yershov G、Dubiley S等以β珠蛋白为研究对象，探讨基因芯片对地中海贫血诊断的价值。1996年，Gennady yershov同时采用杂交测序方式与邻堆杂交阵列方式设置基因芯片，对地中海贫血的基因突变进行检测，结果发现两种方式均能有效地诊断IVS－1－1、IVS－1－5、IVS－1－6等突变，还能有效地区分纯合子与杂合子突变，最近，中国深圳赛尔公司成功地推出了Thalachip地中海贫血基因诊断芯片，采用针对已知突变设计相应的寡核苷酸探针，同时将检测信号与正常序列进行比较，能一次性检测中国人地中海贫血常见的29种突变类型，可覆盖99%以上的突变人群。突变位点包括IVS1－5(G－C)，IVS1－1(G－C)，HbE26(G－A)，β28(＋C)，17(A－T)，－28(A－G)，－29(A－G)，－30(T－C)，14－15(＋G)，1(T－G)，β71－72(＋A或＋T)，β41－42(－TCTT)，42(＋T)，43(G－T)，IVS2－1(G－T)，IVS2－654(C－T)，IVS(2－5)，95(＋A)等29种。

2.葡萄糖－6－磷酸脱氢酶缺陷症基因突变检测芯片。葡萄糖－6－磷酸脱氢酶(G6PD)缺陷症(glucose－6－phosphate dehydrogenase deficiency)属X连锁不完全显性遗传病，由红细胞内G6PD遗传性缺陷所致。可因药物、蚕豆或其他因素而诱发溶血性贫血。G6PD缺陷者在人群中分布很广，多见于黑人、地中海地区及东方民族。中国广东、广西、四川、福建等地区的发现率较高，广东地区约为8%，广西壮族自治区约为15.6%。在临床上主要表现为蚕豆病、药物性溶血、新生儿黄疸及慢性非球形细胞溶血性贫血等。

G6PD基因位于Xq28，由13个外显子和12个内含子组成，目前已鉴定的G6PD生化变异

型有 400 余种，DNA 突变型已发现 100 多种，在中国有近 20 种，G6PD 基因突变型具有下述特点：①多为单个碱基置换的错义突变；②大多数基因突变属于转换型突变，常在 CpG 二核苷酸内发生 C－T 的转换；③G6PD 基因突变具有种族和地区异质性：如在地中海沿岸国家及中东地区 G6PD Mediterranean 常合并 nt 1311C—T 突变，而在亚洲 G6PD Mediterranean 不合并 nt 1311 C—T 突变；④G6PD 基因中还未发现整个基因或大片段基因的缺失，也未见无义突变及移码突变；⑤除外显子 I 外，基因突变遍及其余外显子；⑥G6PD 变异型具有复合突变。如 G6PD A－有三种复合型（nt 376A－G/nt 202G－G，nt 376A－G/nt 680C－G，nt 376A－G/nt 968T－C）。我国人中目前所发现的 G6PD 基因突变型主要如下：

C1（1376G－T）、C2（1388G－A）、C3（1311C－T）、C4（292G－T）、C5（1024C－T）、C6（95A－G）、C7（592C－T）、CT1（835A－T）、CT2（1360C－T）、CT3（493A－C）。

中国上海博星公司结合中国人 G6PD 基因突变的特点，开发出 G6PD 基因突变检测芯片，可以检测中国人最常见的 8 种突变，如 C1（1376G－T）、C2（1388G－A）、C6（95A－G）、C5（1024CT）等，覆盖 90% 以上中国人 G6PD 缺陷的基因突变，是适合国人实际情况的检测手段，经过在广东广西的大量验证工作，目前产品已经成熟，相信不久就可以为提高华南人民群众的健康水平和生活质量贡献力量。

3. 囊性纤维化病基因突变检测芯片。囊性纤维化病（cysticfibrosisdisease，CF）是常染色体隐性遗传病。以西欧、北欧及北美人群发病率为高，约占活产婴儿的 1/2 000，致病基因携带者频率达 1/22，在其他人种和地区发病率则较低。

CF 致病基因定位于 7 号染色体长臂 3 区 1 带（7q31）。全长约 250 kb，共有 27 个外显子，cDNA 全长约 6 129 bp，编码 1 条长 1 480 个氨基酸的肽链，后者被命名为囊性纤维跨膜转运调节物（CFTR），调节通道蛋白的氯离子的转运。迄今发现的 CF 基因突变已达 600 种，除 ΔF508 及另外十余种为较常见的突变外，其余均为罕见突变或仅见于个别病例。已发现突变类型主要有：①密码子缺失，最常见的为 ΔF508，即由于第 10 外显子 3 个碱基对缺失，使肽链第 508 位上苯丙氨酸缺失。这一突变约占突变染色体的 67%；②错义突变是除 ΔF508 外最为常见的突变类型，约占其他突变的一半以上；③移码突变；④无义突变；⑤剪接位点突变等。

Cronin MT 针对囊性纤维跨膜转运调节基因 11 号外显子调节区域的 37 个已知突变，建立囊性纤维跨膜转运调节基因突变检测芯片，采用盲法对 10 个基因组 DNA 进行检测，通过杂交信号的差别能成功地进行基因分型，了解究竟为野生型以及突变是纯合子还是杂合子。Erilich 和 Gelfand 针对胆囊纤维化患者设计的对其跨膜转运调节基因突变进行检测的一种基因芯片，已于 1998 年上市，对研究该病的发病机理、诊断以及预后具有重要的价值。

4. 酪氨酸酶基因突变检测芯片。白化病（albinism）是 Garrod 于 20 世纪初开始研究的遗传代谢病之一，发病率为 1/10 000～1/20 000。可分眼皮肤白化病Ⅰ、眼皮肤白化病Ⅱ和眼白化病 3 大类。不同类型的白化病，临床表现不同，共同的特点是黑色素细胞的酪氨酸酶缺乏，或酪氨酸酶相关酶异常的共同作用，使黑色素小体内酪氨酸不能转化为黑色素，造成毛发、皮肤、眼的颜色变浅，故名白化病。

眼皮肤白化病Ⅰ多为常染色体隐性遗传，酪氨酸酶基因位于 11q14·q21。该基因含 5 个外显子，总长度＞50 kb。已发现 36 种突变导致酪氨酸酶活性消失，引起眼皮肤白化病Ⅰ。眼皮肤白化病Ⅱ型也称酪氨酸酶阳性眼皮肤白化病，是最常见的白化病类型，基因定位于 15q11.2～q12，也称 *P* 基因，已在此型患者中发现该基因的突变导致酪氨酸酶活性消失，主要为 X 连锁隐性遗传（基因定位在 Xq26.3－27.1）。

1994 年，Guo 等根据酪氨酸酶基因第四外显子区域的 5 个单碱基突变的特点，用玻璃作为基片，设计了针对突变的等位基因特异性寡核苷酸(ASO)微阵列，建立了简单快速的基因突变的分析方法。将 ASO 共价固定于玻璃载片上，采用 PCR 扩增基因组 DNA，其一条引物用荧光素标记，另一条引物用生物素标记，分离两条互补的 DNA 链，将荧光素标记 DNA 链与微阵列杂交，通过荧光扫描检测杂交模式，即可测定 PCR 产物存在的多种突变，该方法对人的酪氨酸酶基因第 4 个外显子内含有的 5 个单碱基突变进行分析，结果显示单碱基错配与完全匹配的杂交模式非常易于区别，可快速、定量地获得基因信息。

5.线粒体基因病基因芯片。线粒体是细胞内的一个重要细胞器，存在于人体所有的细胞内，是细胞的氧化中心和动力站，不同的组织细胞中线粒体数目有所差异，如肝脏、心肌、骨骼肌等组织细胞中线粒体数目含量较多，而精子中则较少；每个线粒体中又有多个拷贝的线粒体 DNA(mitochondrial DNA，mtDNA)，因此每个细胞均含有很多线粒体拷贝。

线粒体基因组由线粒体 DNA(mtDNA)构成，人 mtDNA 长 16 569 bp，是由两条链组成的环状 DNA 分子，根据两条链所含碱基成分不同，分别称为轻链(L 链)和重链(H 链)。线粒体基因组由 37 个基因组成，共编码 2 种 rRNA、22 种 tRNA 以及 13 种多肽，其中 3 种是细胞色素 c 氧化酶复合体的亚单位，2 种是 ATP 酶复合体 F 因子的亚单位，7 种是 NADH－CoQ 还原酶复合体的亚单位，1 种是 $CoQH_2$ 细胞色素 C 还原酶复合体的亚单位。因此 mtDNA 的正常表达是维持细胞氧化代谢的重要因素。

线粒体 DNA 的突变导致细胞的氧化磷酸化过程缺陷，直接影响细胞的呼吸与能量获取，导致相应的遗传性疾病，主要有以下几类：

(1)由点突变引起的疾病，如 Leber 遗传性视神经病(Leber hereditary optic neuropathy，LHON)，mtDNA 上许多位点的突变都会导致 LHON。1988 年 Wallace 最先发现 mtDNA 第 11 778 位点的 C 转换成了 A，使 NADH 脱氢酶亚单位 4(ND4)蛋白质中第 340 个氨基酸由精氨酸变成了组氨酸。随后研究发现大约 50%的 LHON 病是由该位点引起的。其他点突变包括 13 708(G～A)、5244 (C～A)、4917 (A～C)和 4216(T～C)等 10 多种，其他由点突变引起的疾病还有线粒体肌病脑肌病伴乳酸中毒及中风样发作综合征(mitochondrial myopathy，enceohalooth，lacticacidosis，and stroke－like episodes，MELAS)、肌阵挛性癫痫伴碎红纤维病(myoclonic epilepsy and ragged－red fiber disease，MER－RF)、神经性肌软弱、共济失调、色素性视网膜炎(neuropathy，ataxia and retinitis pigmantosa，NARP)等。

(2)mtDNA 缺失引起的疾病，如 Kearns－Savre 综合征，几乎所有的患者均有 mtDNA 缺失。缺失片段大小范围在 2.0～7.0 kb 之间，缺失大多发生在重链与轻链的两个复制起始点之间。由于缺失丢掉了许多基因，尤其是 tRNA 基因，因此受这些缺失突变影响的所有个体都呈现出不同程度的 mtDNA 蛋白质合成缺陷，影响所有 4 种呼吸链复合物。另外，Sehon 还发现大约有 1/3 的 KSS 患者的缺失发生在 4 977 bp 区域，断裂点分别位于 ATP 酶亚单位 8 基因中的第 8 468 位和 ND5 基因中的第 13 446 位；而且从 DNA 序列分析可知，缺失两侧有一个 13 bp(－ACCTCCCTCACCA)的同向重复序列，缺失后的融合基因编码出异常的蛋白质产物。这种两侧重复序列可能因同源重组而产生了缺失。这种缺失在 30%～50%的病人中可以见到。最大片段的缺失则为 15 945 位点到 5 786 位点之间 10 kb 的缺失及 5 448 到 470 位点间的缺失。其他如 Pearson－骨髓/胰腺综合征、线粒体心肌病(mitochondrial cardiomyopathy)、帕金森病以及 Alzheimer 病等。

(3)mtDNA 重复引起的疾病较少。

Chee M 于 1996 年在 science 发表文章，对全长约 16.6 kb 人线粒体基因组基因，采用叠瓦式阵列设计基因芯片，每条探针长约 15 个碱基，共 165 000 条探针，每个探针之间的空间间隔为 35 μm，固定在硅化的玻璃片上形成基因芯片，对人类线粒体基因组进行测序，发现其准确率高达 99%，而且从 11 个非洲人线粒体基因组中检测出 505 个多态性位点，并在 Leber 遗传性视神经病 mtDNA 上检测出三个突变位点。

6. *P*53 基因突变检测芯片。*P*53 基因定位于人类染色体 17q12－13.3，在染色体基因组中跨距 20kb 的区域，由 11 个外显子组成。它的 mRNA 长 2.5 kb，编码一个由 393 个氨基酸组成的蛋白质，蛋白质的分子量为 53 kDa，*P*53 因此而得名。野生型 *P*53 基因为抑癌基因，可以使细胞停留在 G1 期，也可以启动凋亡程序和细胞死亡程序。因而，能防止机体中有遗传损伤的细胞的繁殖，降低肿瘤克隆系出现的风险。一旦发生突变，*P*53 基因可能丧失抑癌基因，而且有可能协助 *Ras* 基因，成为癌基因。因此检测 *P*53 基因对研究肿瘤的发生与发展以及预后具有重要的价值。

Drmanac S 等根据杂交测序的原则，设计了一含 8 192 个探针的基因芯片，对 *P*53 基因 5～8 外显子，约 1.1 kb 长的片段的突变进行检测，能准确地检测 12 份样本中存在的碱基的替代、插入或缺失突变，还能区分为纯合子和杂合子。美国纽约 Roswell Park 肿瘤研究所则推出成型的 *P*53 基因芯片，可用于检测 *P*53 的基因突变与基因多态性，采用类似叠瓦阵列的方式设计探针阵列，只不过在叠瓦阵列增加一条探针，该探针不含突变位点对应的核苷酸，通过荧光强度的强弱反映是否存在基因突变，野生型的靶序列具有最强的荧光信号，Wen WH 等人分别用 DNA 测序分析和基因芯片两种方法识别 108 例卵巢癌肿瘤样本中 *P*53 突变，结果所有基因芯片识别的突变位点均能被传统的 DNA 测序所证实，检测准确率为 94%，灵敏度 92%，特异性 100%，证明基因芯片是一种相当有效的突变检测手段，而且发现有 *P*53 突变的病人其生存时间明显低于没有突变的患者。中国上海博星公司推出了有自主知识产权的 *P*53 基因突变检测芯片，可对该基因进行长期监控，尽可能早期发现体内产生肿瘤的迹象，检测位点为密码子 157，175，245，248，249，273，282 的突变。

7. *Brca* 1 基因突变检测芯片。*Brca* 1 为抑癌基因，位于人类染色体 17q21，长约 100 kb，由 24 个外显子组成，乳腺和卵巢细胞若缺少 *Brca* 1 基因表达的产物便会生长失控，导致肿瘤的发生。目前发现 *Brca* 1 基因的突变与遗传性乳腺癌和卵巢癌的发生密切相关。

基因芯片技术可快速、准确地检测出 *Brca* 1 基因的突变与多态性，1996 年 Hacia 等人使用高密度基因芯片，包括 96 000 个 20 mers 的 ONG 探针，检测遗传性乳腺癌与卵巢癌患者 *Brca* 1 第 11 个外显子、全长 2.45 kb 的所有可能的突变，在 15 例已知变异的样品中 14 例发现突变，20 个对照没有出现假阳性结果，检测准确率达 99%，同时检测出 8 个单核苷酸多态性位点。2000 年 Reyna Favis 等报道采用通用的 DNA 阵列对 *Brca* 1 185 del AG 、*Brca* 1 5382 ins C 以及 *Brca* 2 6174 del T 进行检测，结果发现该阵列能有效地区分这些小片段的插入或缺失引起的突变。

8. 其他。尽管目前遗传性疾病基因芯片的数量有限，而且只是针对单一遗传病的部分或大部分突变，但是随着基因芯片技术的成熟以及疾病基因组计划的实施，人类遗传病的致病基因及其相关基因的明确，因此根据各种遗传病的特点，设计出相应的基因芯片甚至多种疾病的联合诊断芯片，应在不久的将来得到实现。

五、遗传性疾病的基因芯片的应用

(一)检测基因突变与多态性,用于遗传性疾病的普查和产前检查等

随着人类基因组草图的完成,许多遗传性疾病的相关基因被相继定位,逐步建立遗传病的疾病基因组学,为从基因水平认识遗传病并进行早期诊断奠定了基础,如地中海贫血、苯丙酮尿症、血友病、以及葡萄糖-6-磷酸脱氢酶(G6PD)缺陷症等。而基因芯片技术针对不同的基因突变或基因多态性,设计出不同的寡核苷酸探针,并固定在固相载体上,即基因芯片,同时采用 PCR 技术对含突变或多态性位点的基因片段进行扩增,将扩增片段与基因芯片进行杂交。从某种意义上,基因芯片技术是 PCR/ASO 技术的集合,是大通量、高容量的 PCR/ASO 技术,能同时对成百上千种基因突变或多态性位点进行检测,这是其他任何基因诊断技术所不能比拟的,因此基因芯片技术,是进行大规模遗传病基因突变检测的最有效手段,如 Guo 等人采用基因芯片,对人酪氨酸酶基因第 4 外显子内含有的 5 个单碱基突变进行分析,结果显示单碱基错配与完全匹配的杂交模式非常易于区别,可快速、定量地获得基因突变信息,而我国深圳赛尔公司成功地推出了 Thalachip 地中海贫血基因诊断芯片,能一次性检测中国人地中海贫血常见的 29 种突变类型,覆盖 99% 以上的中国地中海贫血突变人群。

众所周知,遗传病是一类严重危害人体健康的疾病,许多遗传病患儿一出生即为先天畸形、智力障碍,或长期需要药物治疗,这严重降低了人民的生活质量,因此检测遗传病不是医学工作者的最终目的,其最终目的应为预防遗传病患儿的出生,而产前诊断是预防遗传病患儿出生最有效的手段,长期以来通过采集羊水或胎儿血细胞进行产前诊断,具有较高的危险性。最近,有研究者建议从母体血液中分离出胎儿有核红细胞,采用基因芯片技术,检测胎儿单个有核红细胞中是否存在基因突变,用于遗传病的产前诊断,降低了产前诊断的风险,易于被广大父母亲所接受。因此一旦基因芯片技术研究成熟,采用基因芯片技术,就能有效地对多种疾病进行产前诊断,预防遗传病患儿的出生,真正实现我国优生优育的基本国策。

当然,也可采用基因芯片技术进行婚前检查,降低遗传病患儿发生的机率。我们知道,染色体上的等位基因分显、隐性。显性基因能够表达性状,而隐性基因和显性基因在一起时,它的作用就会暂且被埋没,并不表现,称为携带者,只有当两个隐性基因在一起时,才会致病。有许多遗传病的致病基因为隐性基因,多数人只是携带者而不发病,但如果是两个携带者结婚生子,其后代的发病率为 1/4。如果采用基因芯片,在婚前对多种遗传性疾病进行联合检测,可以很方便地查出隐性致病基因的携带者,从而避免两个携带者结婚生子,显著降低遗传病患儿发生的机率。

(二)用于遗传性疾病的基因分型

基因型是与表型相对应的概念,表示个体某一等位基因的组成状况,根据同源染色体某一基因座位上等位基因的变化,可以将基因型分为两大类,一类是纯合子,另一类是杂合子。纯合子指同源染色体同一基因座位上的等位基因是相同的,而杂合子则指同源染色体的等位基因是不同的,如存在于同源染色体的等位基因为 A 和 B,则 AA 为纯合子,AB 则为杂合子;如果两个基因座位上的等位基因均为杂合子,称为双杂合子。对于遗传性疾病,当同源染色体的某一基因座位的等位基因发生突变时,如果等位基因均发生突变,则称为突变的纯合子,如只有等位基因的某一个发生突变,则称为突变的杂合子;和双杂合子不同,将同一基因座位上出现两种不同的突变等位基因又称为复合杂合子。基因芯片技术不仅能够明确是否存在基因突

变，而且能用于遗传性疾病的基因分型，是突变的纯合子还是杂合子。例如根据β地中海贫血IVS－1突变的特点，直接设计与突变序列互补突变探针，与参考探针一起组成一个阵列，具体如下：

1.正常与突变序列：

正常序列：5－TGGTGAGGCCCTGGGCAGGTTGGTAT－3

IVS－1－1：5－TGGTGAGGCCCTGGGCAGaTTGGTAT－3′

IVS－1－5：5－TGGTGAGGCCCTGGGCAGGTTGcTAT－3′

IVS－1－6：5－TGGTGACCCCCTGGGCAGGTTGGcAT－3′

2.探针序列及探针在阵列的位置：

A(参考探针)：3－TCCAACCATA－5′

B(IVS－1－1探针)：3－TCtAACCATA－5′

C(IVS－1－5探针)：3－TCCAACgATA－5′

D(IVS－1－6探针)：3－TCCAACCgTA－5′

3.检测结果：如图17－9(图中"○"表示有特异的荧光信号产生)。

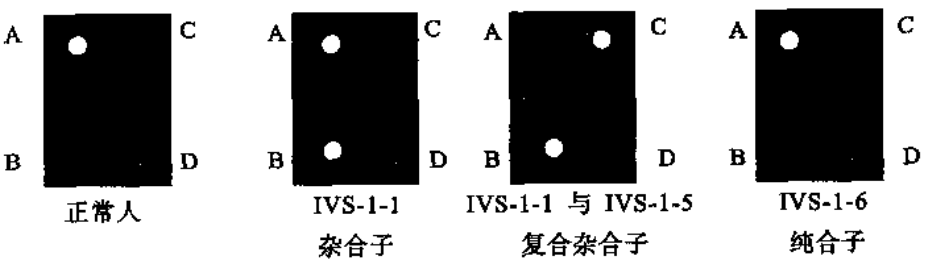


图17－9　微阵列检测β地中海贫血IVS－1突变结果示意图

(三)用于遗传性疾病的基因组物理作图

基因组物理图是用于明确基因位点在染色体上的位置以及其与其他基因的距离的图谱，通常以一段已知核苷酸序列的DNA片段为位标(如STSs，已定位的序列标记位点)，以Mb或kb作为图距的基因组图，在建立物理图时，首先采用机械或限制性内切酶将基因组DNA切成便于操作的大片段，再将大片段克隆到适当的载体中然后通过位标对各克隆进行排序，从而将DNA片段拼成邻接克隆群。目前常用的克隆载体为YAC(酵母人工染色体)和BAC(细菌人工染色体)等，Sapolsky等人采用高密度DNA探针阵列和相应的生化与信息学方法进行基因组文库克隆排序，捕获散在于整个高分子量DNA中分散的短序列标志物，这些标志物可同含互补探针的高密度寡核苷酸阵列杂交，该方法对基因组文库中每一克隆，均可检测出一批特征性序列标志物，而每一对克隆中，在标志物之间相似性基础上，测量其相对重迭部分，利用该重迭信息，即可连续性地将文库排列成重叠群图谱。如采用基因芯片明确STSs在YAC克隆载体的分布情况如表17－5，而其重叠群图谱则如图17－20。

表 17－5　STSs 在 YAC 克隆载体的分布情况

YAC 克隆	STS－A	STS－B	STS－C	STS－D	STS－E	STS－F
1	+	－	－	－	－	－
2	+	－	－	+	－	－
3	－	－	－	+	－	－
4	－	+	－	+	－	－
5	－	+	+	－	－	－
6	－	－	+	－	+	－

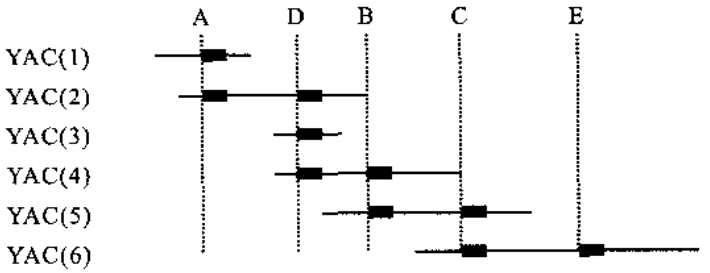


图 17－20　YAC 重叠群图谱

（李一荣　吴健民）

第四节　基因芯片技术在肿瘤研究中的应用

基因芯片技术（DNA chip technique），或称 cDNA 微阵列杂交技术（cDNA miroarray hybridization），可以同时检测分析样本中成千上万个基因的表达变化情况，因此它在肿瘤研究领域中也有着广泛而独到的应用。

一、肿瘤基因表达谱的研究

采用基因芯片技术检测肿瘤基因的表达谱是芯片技术在肿瘤研究中的基础，只有获知整个肿瘤基因组的表达及结构变化，才能从整体上全面了解肿瘤细胞基因型与表型间的关系。肿瘤细胞基因组的改变及其相互作用最终决定了肿瘤的发生发展过程，通过基因芯片技术获得正常细胞和肿瘤细胞在不同阶段不同状态下的表达谱，然后通过生物信息学处理，对这些表达谱进行比较和统计分析，确定正常细胞和肿瘤细胞的基因表达上的差异以及不同基因在表达上的相关性，进而帮助人们在基因水平上深入了解肿瘤的发生发展。

1995 年美国 Stanford 大学的 Schena M 等首次报道了用基因芯片技术定量检测基因的表达，从此基因芯片越来越多地应用于各种基因表达谱和基因变化监测的研究。芯片技术在肿瘤方面的应用，更是极大推动了肿瘤发生发展在基因分子水平的研究。

DeRisi 等研究了人黑色素瘤细胞系 UACC－903 和引入人类正常 6 号染色体的人黑色素瘤细胞系 UACC－903（＋6）的基因表达谱，结果发现了一些先前未认识到的基因变化，这些变化可能导致了两种细胞系的生长特性和裸鼠致瘤性的不同。UACC－903（＋6）和 UACC－903 相比具有显著不同的性质，前者生长速度降低、恢复接触抑制并失去裸鼠致瘤性。作者用含有

1 161个基因片段的芯片与来源于这两种细胞分别标有不同荧光染料的 cDNA 杂交，比较分析两者的基因表达水平的变化，再用 Northern - blot 进行验证，发现有致瘤性细胞的基因表达谱中，单核细胞趋化因子和单核细胞活化因子基因等表达升高，α1 - 抗糜蛋白酶基因等表达降低。

DNA 芯片尤其适用于病因复杂的肿瘤基因表达谱分析情况。例如，恶性肿瘤的发生常常是多基因共同作用的结果，DNA 芯片技术在肿瘤细胞基因表达模式及肿瘤相关基因发掘中具有重要的作用。Wang 等将一些看家基因、细胞因子及受体基因、细胞分裂相关基因及其他一些癌基因共 5 766 个基因的 cDNA 探针固定在芯片上，对正常卵巢组织及卵巢癌组织的 mRNA 进行分析，发现二者之间 30% 基因表达相差 2 倍以上，9% 相差 3 倍以上，其中上调较为明显的有 CD9 分子、上皮糖蛋白(epithelial glycoprotein)、P27 及 HE 蛋白激酶抑制物等。这些结果不仅进一步证实了以前用其他方法获得的结果，还提供了一些新的信息，为卵巢癌的诊断和治疗提供了依据。

Bubendorf L 等用 cDNA 微阵列和组织微阵列研究了前列腺癌激素治疗失败的分子机制。作者先用 cDNA 微阵列检测了抗激素 CWR22R 细胞前列腺癌异种移植物(人前列腺癌移植入裸鼠)和激素敏感 CWR22 细胞前列腺癌异种移植物的基因表达谱，找出表达有差异的基因。然后，用大量临床肿瘤组织和正常对照制成组织芯片，再通过免疫组织化学检测样本中特定蛋白质的表达水平，以确证那些表达有差异的基因，临床组织样本包括 26 例良性前列腺增生组织、208 例原发性前列腺癌组织和 30 例抗激素局部复发性前列腺癌组织。在 5 184 个被检测的基因中，抗激素 CWR22R 细胞前列腺癌异种移植物中有 37 个基因(0.7%)的表达水平比 CWR22 细胞异种移植物高 2 倍以上，有 135 个基因(2.6%)的表达水平降低超过 50%，胰岛素样生长因子结合蛋白 2(IGFBP2)和 27 kDa 的热休克蛋白(HSP27)在 CWR22R 肿瘤中表达最高。组织微阵列免疫组织化学检测结果表明，IGFBP2 蛋白在 100% 临床抗激素局部复发性前列腺癌中高表达，在 36% 原发性前列腺癌中高表达，在良性前列腺增生组织中无一例表达。热休克蛋白 HSP27 在 31% 抗激素局部复发性前列腺癌中高表达，在 5% 原发性前列腺癌中高表达，在良性前列腺增生组织中也无一例表达。

Sgroi DC 等将基因芯片技术和激光捕获微切割技术(laser capture microdissection，LCM)结合起来，检测了来自同一名乳腺癌患者完全正常乳腺细胞、浸润性乳腺癌细胞和转移性乳腺癌细胞内的基因表达水平。LCM 技术可以从含有不同成分的组织中特异性地获取某一特定的同质细胞，如几十个、几百个同类细胞，甚至是某些单一细胞，其原理是将准备好的切片放在倒置显微镜上，再将切片覆盖一层透明乙烯乙酸乙烯酯(EVA)薄膜，选定要切割的区域，并将该区域调至视野中央，然后将从侧面打过来的激光束通过一个透镜和反光镜，使其从切片上方垂直照射下来，其光路与显微镜聚光器的光路共轴，这样光斑正好落在显微镜视野的中心，即要切割的区域。该区域的 EVA 膜受到适当强度的激光照射后，瞬间温度升高并与其下方的细胞相粘连；当将 EVA 膜揭起时，与之相连的细胞也随之从切片上完好地取出来。如此重复多次，即可从不同部位获取所需的同类细胞。随后，将带有细胞的 EVA 膜放入含有蛋白酶 K 的试管内，使细胞与膜分开，并将细胞裂解，获得 DNA、RNA、蛋白质等待测物质。作者在 3 万个被检测的基因中，将表达有差异的基因再用实时定量 PCR 和免疫组织化学等方法加以确证，然后通过比较分析在正常乳腺组织和不同病理状态乳腺癌组织中的基因表达谱，阐明涉及乳腺癌发生发展的分子事件，进而帮助人们深入认识肿瘤的发生、浸润和转移。由于是采用同一个病人研究乳腺癌的发生发展过程，还可以避免个体差异，减少实验误差，从而得到比较准确可靠的实验结果。

Alevizos I 等采用与 Sgroi DC 类似的方法，检测了体内口腔癌组织和正常口腔上皮细胞的基因表达谱。发现有 600 个基因与口腔癌相关，这些基因包括癌基因、肿瘤抑制基因、转录因子基因、转移蛋白基因以及其他一些还未发现的口腔癌相关基因。

Kitahara O 等在研究结肠癌和正常结肠上皮组织基因表达的变化时，先用 LCM 技术从 8 例结肠癌组织及其相应正常结肠上皮组织取得各自同质细胞群，再与含有 9 216 个基因的芯片杂交，分析比较两者基因表达的变化。结果发现在大多数肿瘤中都有 235 个基因表达发生改变，其中 44 个上调，191 个下调。这些变化的基因与细胞的信号转导、物质代谢酶、活性氧簇的产生、细胞周期、有丝分裂和凋亡有关。对 8 例结肠癌组织及另外 12 例结肠癌组织中的 10 个有表化的基因，包括 5 个上调和 5 个下调基因，通过半定量逆转录 - PCR 进一步证实其变化，以此确认芯片实验结果的可靠性。

Rihn BH 等检测了间皮瘤和正常胸膜细胞的基因表达谱，在超过 6 500 个被检基因中，发现了不到 300 个表达有差异的基因，这些基因大多数与大分子稳定性、细胞粘附和细胞识别、细胞迁移和细胞分化有关。对其中几个基因的表达经 RT - PCR 进行检测证实后，可用作人胸膜间皮瘤的诊断标记物。

国内学者在用芯片研究基因的表达谱方面也做了许多的工作，例如中国科学院上海生物化学与细胞生物学研究所等单位，将 14 000 个基因片段点在尼龙膜上制成芯片，然后用来检测了 9 对肝细胞癌患者癌组织和癌旁正常肝组织的基因表达谱。结果显示每一对被检测的基因表达谱中，平均 1 165 个基因表达有变化，但是在一半以上的肝癌组织中，组成性改变的基因有 156 个，其中 72 个基因（包括 30 条新基因）下调，84 个基因（包括 48 条新基因）上调。最令人感兴趣的是，大多数下调基因是由一组肝脏富含转录因子（liver - enriched transcription factors，LETFs）调控的。另有学者分别进行了肺癌基因、人膀胱癌基因、食管癌、高低转移人卵巢癌细胞系基因表达谱的研究，均发现了许多与相应正常组织差异表达的基因。

二、寻找肿瘤相关基因

寻找并克隆新的肿瘤相关基因是肿瘤研究的一个重要方面，某些肿瘤的发生发展常常是由于某一个或某几个肿瘤相关基因发生突变异常激活而导致，因此寻找并克隆肿瘤相关基因对肿瘤的基因诊断和基因治疗具有重要的意义。然而肿瘤是一类数量大而遗传性不显著的多基因、多因素疾病，采用传统的连锁分析和遗传分析等技术来寻找新基因常常显得力不从心。采用基因芯片技术，在寻找肿瘤相关基因方面有着其他方法不可比拟的优势，它样品用量少，实验周期短，自动化程度高，高通量的生物信息处理，便于快速大量地筛选基因，从而大大提高新基因的发现效率。

肿瘤的发生发展是其相关基因异常变化的结果，因此找到异常表达的基因是寻找肿瘤致病基因或相关基因的关键，芯片技术通过检测大量基因的表达水平，很容易就能筛选到这些异常变化的基因，然后从中发掘起关键作用的肿瘤特异基因。

Moch H 等应用芯片技术寻找肾细胞癌的肿瘤标志物，以便用于肾细胞癌临床诊断和预后的预测。作者首先用含有 5 184 个基因的 cDNA 微阵列，检测了肾细胞癌细胞系 CRL - 1933 与正常肾组织之间的基因表达差异，发现在肿瘤组织中表达有显著差异的基因有 89 个，其中一个差异表达的基因编码波形蛋白，为一种细胞质中间丝；然后采用免疫组化方法，以波形蛋白的抗体为探针来检测 532 个肾细胞癌标本的组织微阵列的表达情况，发现波形蛋白常见于肾细胞（51%）和乳头肾细胞癌（61%），少见于嫌色细胞肾细胞癌（4%）和大嗜酸粒细胞瘤（12%）；同时还发现，

无论患者的疾病分期和病理分化如何,波形蛋白均与其不良预后有明显关系。

Wolf M 等应用基因芯片技术研究了 3 种人骨肉瘤细胞系与人正常成骨细胞基因表达谱之间的差异。在所检测的 5 184 个基因序列中,有 35 个基因表达异常,其中 8 个基因表达升高,27 个表达降低。上调的基因包括热休克蛋白 HSP90β 和多聚腺苷酸结合蛋白样蛋白 1 基因,下调基因包括纤维结合素 1 和血小板反应蛋白 1 基因。对于这些基因的变化,用这 3 种细胞系和与之相应的 3 种主要骨肉瘤样本再通过 RT - PCR 来加以验证。研究结果表明骨肉瘤细胞系的基因表达谱不同于正常成骨细胞,热休克蛋白 HSP90β 和多聚腺苷酸结合蛋白样蛋白 1 基因的上调,与纤维结合素 1 和血小板反应蛋白 1 基因的下调在骨肉瘤的发生发展中都起着重要作用。研究结果还表明骨肉瘤基因表达谱也可用于建立骨肉瘤的分子指纹(molecular fingerprint),应用于患者的临床诊断或从中发现新的基因变化和肿瘤相关分子。

Srivastava M 等用芯片技术证实了抑癌基因 *ANX7* 与前列腺癌的相关性。*ANX7* 基因编码一种 Ca^{2+} 活化的 GTP 酶,参与细胞的外分泌和生长调控,位于人染色体 10q21,这个染色体区域据推测可能存在一些针对前列腺癌和其他肿瘤的抑癌基因。经过实验证明,在 *ANX7* 基因敲除的小鼠中,肿瘤的发生频率明显增加。为进一步验证 *ANX7* 基因是否与前列腺癌有关,作者将野生型的 *ANX7* 基因转入两株前列腺癌细胞 LNCaP 和 DU145 中去,结果发现肿瘤细胞的增殖速度和克隆群的形成显著减慢。再用由 301 份前列腺癌组织样本制成的组织芯片检测 *ANX7* 基因的表达和分布,结果表明与原发性前列腺癌相比,ANX7 蛋白的表达水平在转移性和局部反复发作激素抗性的前列腺癌中呈现明显的减少。另外,对 20 例前列腺癌组织进行杂合子丢失(loss of heterozygosity,LOH)检测,结果 7 例(占 35%)显示存在 LOH。LOH 意味着一个染色体位点上,两个多态性的等位基因中的一个出现突变或缺失,是肿瘤细胞一种常见的 DNA 变异,抑癌基因的 LOH 常导致肿瘤的发生。因此,所有实验结果都有力地证明 *ANX7* 基因具有抑癌基因的生物学和遗传学特性,在前列腺癌的进程中发挥作用。

将基因芯片技术与其他生物技术相结合,尤其是与一些差异筛选技术相结合,用于肿瘤组织中特异表达基因或异常变化基因的检测,一方面可大大增强基因芯片对低拷贝和小片段基因的鉴别能力,另一方面又可省去繁琐费时的大规模基因克隆,减少工作量。比较常用的差异筛选技术是消减杂交法、代表性差异分析技术和抑制消减杂交法等,另外,还有学者用到差异免疫吸附和基因表达系列分析等技术。

经典的消减杂交法(subtractive hybridization,SH)是将实验组 mRNA 与对照组 cDNA 杂交,或将实验组 cDNA 与对照组 mRNA 杂交,去除杂交体和未杂交上的对照组成分 (cDNA 或 mRNA),即可得到实验组独有的 mRNA 或 cDNA,这些基因就是差异表达的基因。Xu J 等将前列腺癌组织和正常人前列腺组织进行消减杂交,得到富集了的前列腺癌组织特异性的 cDNA 消减库。将高丰度的克隆通过菌落杂交而去除,剩下的克隆经 PCR 扩增后点样制成 cDNA 微阵列。用不同荧光素分别标记来自于前列腺癌组织和正常人前列腺组织的 cDNA,混合后与微阵列杂交,挑选出表达有显著差异的基因进行测序。对 3 条新发现的前列腺癌相关基因进一步用 Northern 杂交、实时 PCR 和免疫组织化学等方法鉴定,并克隆到全长 cDNA。生物信息学预测其翻译产物的亚细胞定位均为膜蛋白。后来,该研究小组对其中一条新发现的命名为 *P504S* 的基因作了深入研究,结果表明 *P504S* 在前列腺癌组织细胞中能产生很强、很特异的胞浆染色颗粒,可作为前列腺癌高度灵敏高度特异的阳性标志。此外,该研究小组采用包括芯片技术在内的一整套基因组学研究方法对前列腺及其癌变组织进行研究,发现了一条新的只在前列腺组织细胞中特异表达的基因,该基因的表达与前列腺组织是否癌变及癌变程度无关,可作为前列腺组织

或细胞特异的分子标记，这对前列腺癌的诊断和治疗也具有潜在的应用价值。

Villaret DB 等把 SH 和基因芯片用于寻找头颈鳞状细胞癌(head and neck squamous cell cancer，HNSCC)相关基因。将来自 2 例新鲜冻存的 HNSCC 组织和 6 例正常对照组织进行消减杂交，消减杂交库中的克隆经 PCR 扩增后制成芯片。芯片与来自 16 例 HNSCC 组织和 22 例正常对照组织的 cDNA 探针杂交，再经测序后发现 107 个基因表达有差异，对 13 个在头颈鳞状细胞癌过度表达的基因进行分析，发现其中 9 个为已知基因，4 个为新基因。

传统的消减杂交法对 mRNA 的质和量要求都很高，而且还很难获得低丰度的差异基因。为此，人们引入了 PCR 技术，先后建立了几种以 PCR 为基础的消减杂交方法，代表性差异分析技术(representational difference analysis，RDA)即是其中一种。RDA 的原理是首先用限制性内切酶分别消化两组双链 cDNA，两端接上单链接头 R，补平后用含接头 R 序列的引物扩增，再用内切酶去除双链 cDNA 两端的接头 R；然后单给实验组双链 cDNA 两端再接上单链接头 J，实际上接头是加在 cDNA 每条链的 3′端。将实验组与对照组 cDNA 杂交，杂交反应体系中只有实验组自身杂交形成的双链 cDNA 才两端都带接头 J，补平后可被含有接头 J 序列的引物几何扩增，而只有一端带接头 J 的杂交体只能被线性扩增，并可被大豆核酸酶去除。将此杂交产物再行第二轮酶切、加接头、杂交和 PCR 扩增，重复若干次后，可确保从实验组中彻底去除与对照组共有的序列。Welford SM 等将 RDA 和芯片技术相结合，检测了在 Ewing 肉瘤组织中特异表达的基因。作者对两例在体内具有不同生物学活性的原发性 Ewing 肉瘤进行研究，移行和进展快，另一例表现为局限性并成功治愈，经 RDA 消减杂交获得消减文库，将消减文库用 PCR 扩增后点样制成 cDNA 微阵列。分别用 Cy3 和 Cy5 标记这两种表现各异肿瘤组织的 cDNA，混合后与 cDNA 微阵列杂交，然后检测各点的荧光强度比值。在 192 个被检测的基因中，有 173 个基因的表达存在 2 倍以上的差异。为评估这些基因是否不同及不同程度，进一步将 192 基因进行测序和回杂交(back－hybridization)，结果发现 50 个基因是特异表达的。

抑制消减杂交法(suppression subtractive hybridization，SSH)是另外一种基于 PCR 的消减杂交技术，其原理与 RDA 有些类似，也是先将双链 cDNA 酶切为短片段，但后续工作是将实验组分为两组加不同的单链接头（这些接头实际上是连于 DNA 每条链的 3′端)，分别为接头 1 和接头 2，它们分别具有一段反向末端重复序列。将实验组的两组样品分别与对照组过量的酶切产物杂交，第一轮杂交后，两组杂交产物混合，将这些产物再与对照组进行第二轮杂交，杂交产物补平后作为模板，用含反向末端重复序列的引物进行 PCR 扩增。杂交产物中对照组自身杂交产物和实验组与对照组的杂交体因末端不带接头或只有一端有接头而不能被扩增。杂交产物中还有三种两端带接头的产物，它们都是实验组 cDNA：A 产物为两端带接头 1 的双链 cDNA，每条单链 cDNA 来源于第一次杂交的同一体系；B 产物为两端带接头 2 的双链 cDNA，每条单链 cDNA 来源于第一次杂交的同一体系；C 产物为两端分别带接头 1 和 2 的双链 cDNA，每条单链 cDNA 来源于第一次杂交的不同体系，更具有特异性，而 A 产物和 B 产物带含反向末端重复序列的相同接头，所以在 PCR 变性和复性过程中，每条单链 cDNA 可自身头尾互补结合形成发卡结构而不能被扩增，只有 C 产物才能被扩增出。Yang GP 等将两株分别为雌激素受体(ER)阳性和雌激素受体(ER)阴性的乳腺癌细胞系进行抑制消减杂交，获得 322 个消减杂交子，经 PCR 扩增后点制成 cDNA 微阵列。再用不同荧光素分别标记来自于 ER 阳性和 ER 阴性乳腺癌细胞系的 cDNA，将制成的探针混合后与微阵列杂交，筛选出 10 个在 ER 阳性乳腺癌细胞中表达比对照至少高 5 倍的基因，这些过表达基因再用 Northern 杂交证实。测序分析表明，其中 4 个基因分别为细胞角蛋白 19 片段、*GATA*－3、*CD24* 和谷胱甘肽－*S*－转移酶 *mu*－3 基

因，另 4 个基因是来自于两个不同基因的 ESTs，还剩下的 2 个基因为新基因。最后，用流式细胞术和免疫荧光分析表明 CD24 蛋白在 ER 阳性乳腺癌细胞中过表达。

Bangur CS 等将 SSH 和芯片技术用于研究小细胞肺癌（small cell lung carcinoma，SCLC）的特异表达基因。将含 2 400 个基因的 cDNA 芯片与 32 对各种类型 SCLC 和正常对照样本进行杂交分析，筛选出 209 个上调基因，对其中 43 个基因用实时定量 PCR 验证，并进一步研究其中 9 个特异性可能最强的基因，发现 7 个为已知基因，2 个为新基因。

Liau LM 等在研究恶性胶质瘤的免疫治疗时，将差异免疫吸附（differential immuno - absorption，DIA）用于恶性胶质母细胞瘤（glioblastomas，GBM）特异性抗原及特异基因的筛查，从而建立了另一种寻找肿瘤特异基因的方法。DIA 的原理如图 17 - 21 所示，首先分别制备出 GBM 和正常脑组织的匀浆抽提液作为抗原，然后将两者分别与琼脂糖交联制成亲和层析柱备用。用 GBM 抗原免疫兔子制备抗血清，抗血清过交联有 GBM 抗原的亲和层析柱，得到能与 GBM 抗原结合的抗体。将此抗体再过交联有正常脑组织抗原的亲和层析柱，分离出与正常脑组织抗原有交叉吸附的抗体后，未能与正常脑组织抗原结合的抗体即为 GBM 的特异性抗体。用此特异性抗体筛选 GBM 的 cDNA 噬菌体表达文库，即可筛选到在 GBM 中特异表达的基因。将筛选到的 GBM 特异噬菌体阳性克隆片段亚克隆入质粒载体，插入片段再经 PCR 扩增后制成 cDNA 微阵列，然后用不同荧光素标记的 GBM 和正常脑组织的 cDNA 与此微阵列杂交，根据芯片杂交数据从中找到 GBM 的特异相关基因。实验结果显示，经 DIA 共筛选出 28 个克隆，其中 9 个为已知基因，19 个为新基因。用 cDNA 微阵列杂交检测这些基因在多种不同人胶质瘤中的表达，发现一条基因在肿瘤组织中持续高表达，比正常脑组织可高 3 ~ 30 倍。进一步研究表明，该基因为颗粒酶蛋白基因，与胶质瘤的生长调控有关，在神经胶质肿瘤中高表达，在正常脑组织和其他肿瘤中不表达或低表达。

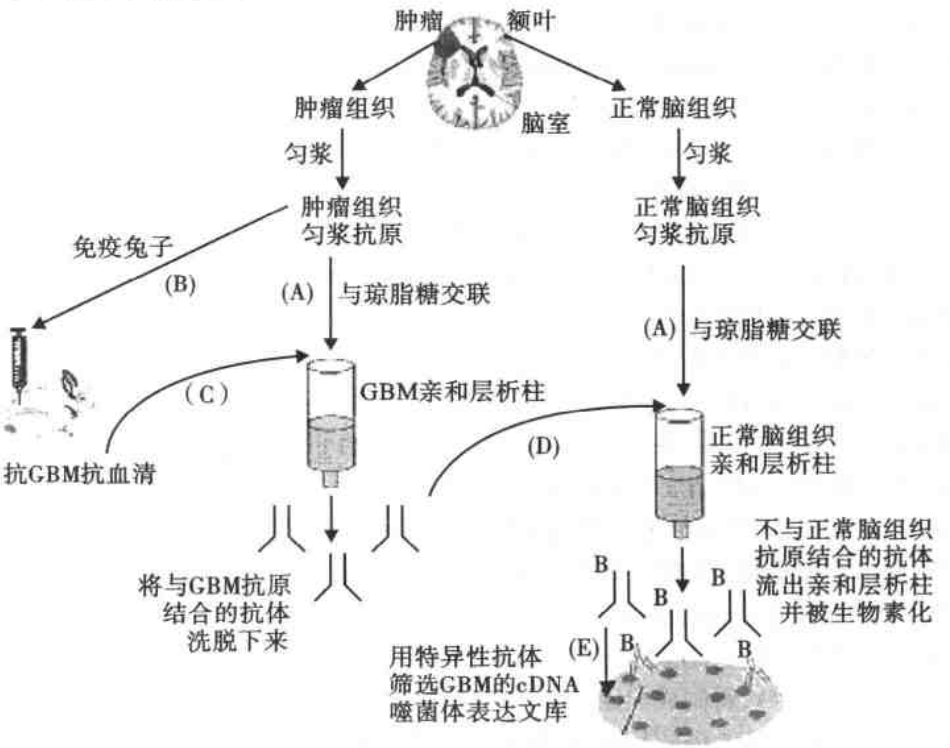


图 17 - 21　差异免疫吸附 DIA 的原理

图 17－21 显示原理说明：①GBM 肿瘤组织和正常脑组织在盐溶液中匀浆后倒出可溶部分，不溶物重新用 1% SDS 抽提，抽提物分别与 CNBr 活化的琼脂糖交联制成亲和层析柱。②将剩余的溶于盐溶液中的 GBM 肿瘤组织匀浆液用完全福氏佐剂乳化后免疫兔子；SDS 抽提的 GBM 肿瘤抗原与不完全福氏佐剂混匀后加强免疫兔子，以促进抗 GBM 抗血清的产生。③抗 GBM 抗血清过交联有 GBM 抗原的亲和层析柱，洗脱后可得到能与 GBM 抗原结合的抗体。④将③得到的抗体再过交联有正常脑组织抗原的亲和层析柱，分离出与正常脑组织抗原有交叉吸附的抗体后，未能与正常脑组织抗原结合的抗体即为 GBM 的特异性抗体。⑤将此特异性抗体经中性化、浓缩、透析和生物素化后用于筛选 GBM 的 cDNA 噬菌体表达文库，即可筛选到在 GBM 中特异表达的基因。

基因表达系列分析（serial analysis of gene expression，SAGE）是同时定量分析大量转录本的一种方法，基本原理是从转录本特定位置（靠近 3′端）分离出短核苷酸序列（9～10bp），它包含了能代表该转录本的足够信息。将一个体系的所有转录本中这一短序列分离并随机连接到一个克隆载体中进行测序，便可以得到该体系所有转录本的表达情况。SAGE 最显著的特点是可用于寻找那些较低丰度的转录物，最大限度地收集基因组的基因表达信息，这使之成为从总体上全面研究基因表达、构建基因表达图谱的一种重要策略。SAGE 使用灵活、便于掌握，在正常、癌旁、癌组织中基因的差异表达研究方面有独到的优点，有助于发现肿瘤特异基因。Nacht M等将 SAGE 和芯片技术结合起来，用于乳腺癌中差异表达基因的检测。首先用 SAGE 对两株典型的乳腺癌细胞 21PT 和 21MT 进行检测，其结果与正常乳腺上皮细胞的 SAGE 数据比较，找出表达有差异的基因和 ESTs。然后，将这些表达有差异的基因和 ESTs 制成点阵，与临床乳腺癌样本和正常对照杂交，以验证 SAGE 的结果。实验数据表明，SAGE 和芯片结果是一致的，除发现一些已知乳腺癌分子标记，如 *HER－2/neu* 和*MUC－1* 在肿瘤组织和正常对照中差异表达外，还发现了一些新的乳腺癌相关基因。Christine A 等在研究胰腺癌相关基因时，先用含 60 000 cDNA 片段（包括 12 000 个全长基因和 48 000 个 ESTs）的芯片检测胰腺癌及其正常对照间差异表达的基因，然后用 SAGE 检测两株正常胰腺细胞，以排除一些在正常胰腺细胞中表达的基因。挑选出差异表达的基因后，再用免疫组织化学、荧光原位杂交和逆转录－PCR 加以证实。最后，在鉴别出的 97 个胰腺癌相关基因中，发现 28 个为已知的，69 个为新基因。

另外，对芯片数据进行合理的统计分析有助于肿瘤相关基因的发现。Thomas JG 等通过研究两组已确证存在差别的样本基因表达谱数据，建立了一套统计回归模型来帮助发现表达有差异的基因。这个模型完全基于假说，应用严格的统计方法，分析各种不同的复杂的基因表达数据。与聚类分析不同，该方法并不侧重于找出众多基因或样品的共同特征，而是侧重于灵敏有效地发现有差异表达的基因。另外，这种方法可用于检验基因表达的统计假说。用这种方法重新分析 Golub TR 等 11 例急性淋巴细胞白血病（ALL）和 27 例急性粒细胞白血病（AML）的基因表达谱数据，发现 141 个基因表达有差异，约占两者相关基因的 1% 。用这种方法比较 AML 组不同样本间的差异，从中发现一组与血小板生成素相关的基因，该基因与 AML 的治疗效果有关。

三、肿瘤分型的研究

目前肿瘤的分类主要还是依靠临床指征和组织病理学镜检，但这种方法有着许多局限，一些组织病理学表现相似的肿瘤在临床转归和治疗效果上可有很大的不同。这迫切需要人们找到一种能更准确区分肿瘤类型的方法或规则，肿瘤细胞的基因变化决定着其功能和性状的改

变,用肿瘤细胞的基因型指标进行肿瘤的分型可以更好地区分辨别肿瘤的细微差异,而基因芯片技术为基于基因表达类型的肿瘤分型方法提供了技术基础。

1999 年,美国麻省理工学院 Whitehead 研究所 Golub TR 等最早利用基因芯片技术进行了肿瘤分型的尝试,在人类急性白血病亚型的分型上获得成功。作者用点有 6 817 条基因的芯片分别与 38 例急性白血病患者骨髓细胞的 cDNA 进行杂交,其中 27 例为急性粒细胞白血病(AML),11 例为急性淋巴细胞白血病(ALL),然后比较分析两者基因表达谱的异同,以此作为区分这两种不同型别白血病的依据。通过他们自己建立的"Neighborhood Analysis"方法对基因表达谱数据进行分析,确定大约有 1 100 个基因与 AML 和 ALL 有相关性,然后又根据其他大量临床检测数据建立了 AML 和 ALL 的相关基因表达数据库,确定各相关基因的平均表达水平作为 AML 和 ALL 分型的参照值。实际应用时,将待查患者的 cDNA 样品与基因芯片杂交,杂交结果与建立的 AML 和 ALL 的相关基因表达数据库及分型参照值进行比较分析,确定它们与 AML 和 ALL 的相关性及相关程度,从而判断病人是否患有急性白血病或判断病人所患急性白血病为何种亚型,所有的数据处理及判断均由计算机自动完成。鉴于与 AML 和 ALL 的相关基因有 1 100 条之多,给实际操作和数据处理带来许多繁琐和不便,作者又从中精简出相关程度更高的 50 条基因作为分型判断的依据。

为验证方法的准确性和可行性,作者用此方法对上述 38 例已明确诊断的急性白血病进行分型,结果计算机成功地对 36 例进行了分型,只有 2 例不能判断。另外又对来自对照实验室的 34 例急性白血病不同组织标本,包括 24 例骨髓细胞和 10 例外周血细胞,进行分型判断,结果有 29 例诊断正确。在随后的实验中,作者又用此方法对 15 例 AML 的化疗疗效及预后进行了分析,研究 AML 的相关基因是否与其发病机制和治疗效果有关。结果在有疗效的一组中没有发现与疗效有显著相关性的基因,在无疗效的一组中发现一种同源框基因(Homeobox gene) *HOXA9* 高表达,而该基因的染色体异位常常出现在一种罕见的预后不良的 AML 亚型中,此外发现该基因高表达可导致体外粒细胞恶性转化和实验动物白血病。

最令人们感兴趣的是,Golub 等用此方法纠正了一个临床误诊病例。一患者临床上表现出典型的白血病症状,但其骨髓标本不具有典型的细胞形态,只好根据临床指征和其他相关检测结果将其诊断为 AML,但按 AML 治疗数月后,病人未见好转,于是怀疑诊断有误。重新将病人的骨髓细胞标本按 Golub 的方法进行芯片杂交处理,结果 AML 和 ALL 的相关基因均表达较低,计算机不能将其归入 AML 或 ALL,从而排除患者是 AML 的可能性。后来,进一步分析基因表达数据,发现样本中腺泡状横纹肌肉瘤相关基因高表达,从而诊断为平滑肌肉瘤,改变治疗方案后病人很快出现好转。这一例证极大地显示了基因芯片技术在肿瘤分型和肿瘤诊断上的应用价值。

Alizadeh AA 等将芯片技术用于了弥散性大 B 细胞淋巴瘤(diffuse large B - cell lymphoma, DLBCL)的分型,分型结果较好地反映了肿瘤的临床表现。DLBCL 是一种侵袭性的恶性成熟 B 细胞淋巴瘤,患者的临床表现常常各异,尽管大多数病刚开始治疗可能会有一定反应,但很少会有病人能得到稳定的病情缓解。以往依靠形态学的诊断,以及免疫组织化学对少数几个诊断标记的检测,都不能有效地将 DLBCL 按其临床表现进行分型。用含 17 856 个基因的芯片,检测包括 DLBCL、滤泡型淋巴瘤(follicular lymphoma, FL)和慢性淋巴细胞白血病(chronic lymphocytic leukaemia, CLL)和正常对照组织等在内的 96 例样本,根据基因的表达水平对 DLBCL 进行分型。结果可将 DLBCL 分为两种亚型,一种为胚胎性中枢 B 细胞样 DLBCL(germinal centre B - like DLBCL, GC - DLBCL),另一种为活化 B 细胞样 DLBCL(activated B - like DLBCL, A - DLBCL),

这两个亚型的基因表达谱显著不同，反映了 B 细胞不同的分化阶段。两个亚型患者的预后明显不同，GC - DLBCL 患者总的生存率明显高于 A - DLBCL 患者的生存率，证实了基因分型结果的准确性和临床意义。

Sorlie T 等根据乳腺癌组织的基因表达特征进行乳腺癌的分型，并探讨了这种基因分型与肿瘤临床预后的相关性，获得了一些新的发现。在对总共 85 例样本，包括 78 例乳腺癌、3 例纤维腺瘤和 4 例正常对照，用含 8 102 个基因的芯片进行杂交分析，根据基因表达特征将乳腺癌分为 5～6 个型别。其中，划分出的基底上皮样组、*ERBB2* 基因过表达组和正常乳腺样组与以前研究的结果一致，而以前划分为腔状上皮/雌激素受体组的样本，这次又可被划分为至少两个亚型，每个亚型都具有独特的基因表达特征。为证明这种新的分型的合理性，对用两套不同基因在肿瘤组织的表达水平进行肿瘤样本分型的结果加以比较，发现两者的分型结果是一致的。两套不同的基因中，其中一套为作者从 8 102 个筛查基因挑选出来的可特异反映各型肿瘤本质的 456 个基因；另一套为挑选出的与患者预后高度相关的 264 个基因，其中有 81 个基因与前一套重复。进一步进行乳腺癌疗效的前瞻性调查，用以评估基因分型的合理性，结果给予同样的治疗以后，不同型别的乳腺癌患者，其预后存在很大的不同，这也进一步证实了基因分型具有其合理性和临床意义。

Selaru FM 等将芯片技术和生物信息学用于巴雷特食管（Barrett's esophagus）和食管癌的分类，结果与组织病理学分类结果一致。用含 8 000 个基因的芯片与 13 例食管外科切除或内镜活组织检查的 RNA 杂交，其中 7 例为巴雷特食管样本，6 例为食管癌，然后用分层聚类分析（hierarchical cluster analysis）进行分型。

Selaru FM 等还将人工神经网络（artificial neural networks，ANNs）用于芯片数据的处理，并进行肿瘤的分型。所谓 ANNs 是指为了模拟生物大脑的结构和功能而构成的一种信息处理系统或计算机系统，由多个非常简单的处理单元彼此按某种方式相互连接而形成，系统靠其状态对外部输入信息的动态响应来处理信息。作者用基于 cDNA 微阵列数据的 ANNs 对两种只有微小差别的肠部肿瘤进行了区分，一种是散在的结肠腺瘤和癌（sporadic colorectal adenomas and cancers，SAC），另一种是炎性肠病相关的发育异常和癌（inflammatory bowel disease - associated dysplasias and cancers），前者只需行息肉切除治疗，后者需要进行结肠切除术治疗。用含 8 064 条基因的 cDNA 点阵与 39 例结肠肿瘤样品杂交，然后根据基因表达数据，分别用分层聚类法（Hierarchical clustering）和 ANNs 进行分类，结果基于所有 8 064 个基因表达数据的分层聚类法无法区别 SACs 和 IBDNs，而 ANNs 却可准确判断出 12 例样本（包括 3 例 IBDNs 和 9 例 SACs）。从 8 064 个数据中再用计算机挑选出 97 个用于分类，结果两种方法都可以对 SACs 和 IBDNs 进行正确分类。

另外，Ben - Dor A 等研究了基因分型的准确性，通过统计分析 3 种肿瘤及其正常对照组织标本的基因表达数据，取得了 90% 的组织分型准确率。第一组数据为 62 例上皮结肠组织的 2 000个基因表达水平，第二组数据为 32 例卵巢组织的约 10 000 个基因表达水平，第三组数据为 72 例骨髓和外周血样品的 7 100 个基因的表达水平，所有数据经过多种统计方法统计并修正后，挑选出表达有差异的相关基因作为分类依据。

可见，基于基因芯片技术的肿瘤分型方法为肿瘤分型提供了一种全新的手段，它与传统的临床指征和组织病理形态学分型方法相比，在某些方面具有更可靠更稳定更准确的优点，更能反映肿瘤组织或细胞的实际状况并预示其发展，实际操作时也更简便快捷，自动化程度高。

四、肿瘤相关基因功能的研究

研究肿瘤相关基因的功能是研究肿瘤致病机制和肿瘤治疗的重要内容，只有明确肿瘤相关基因的功能才能真正在分子水平上认识肿瘤的发生发展，深刻理解肿瘤的本质。

目前，我们对基因功能的研究还主要是靠基因转移或基因敲除手段，通过将待研究的基因逐个转入培养细胞或转基因动物后，从产生的生理或病理效应中推断基因的功能。其实，这种研究方法具有很大的局限性，因为大到生命，小到细胞都是一个有机整体，任何一个细胞乃至任何一个基因都不是单独发挥作用的，它们往往要与多种其他的细胞或基因协同发挥作用。不同的基因可以有相同的功能，同一基因又可能有多种功能，人为地加入或敲除某一基因后，机体或细胞往往通过代偿途径弥补该基因的功能，从而使机体或细胞的表现不能准确反映该基因的真实功能。现代分子生物学和细胞生物学研究表明，许多涉及细胞生长分化、机体免疫和新陈代谢等生命活动的基因为数众多、紧密排列、功能相同或相似，形成一个基因家族，各个基因是作为一个整体来调节的，在细胞的生长分化、个体发育过程中保持高度的时间、空间和数量的协调性。肿瘤的发生发展是一个复杂的过程，一个肿瘤相关基因功能的发挥可能是与许多其他基因相互作用的结果，因此，只有从细胞、组织器官以及整个机体全面考察才能准确认识它的功能。对于种类繁多的各型肿瘤，找出并研究其特异性的相关基因及其基因功能是一项十分繁重和复杂的工作，这迫切要求我们找到一种高通量的分子水平分析技术。

基因芯片技术具有高通量、自动化、简便快速的特点，它将成千上万个基因或寡核苷酸固定在一块芯片上，可以同时对来源于不同个体、不同组织、不同细胞分化阶段、不同病变、不同处理方式的肿瘤细胞的基因表达状况进行测定，从大量的基因表达数据中找出规律性和特异性的肿瘤基因并确定其功能。

Coller HA 等用基因芯片研究了癌基因 *myc* 在细胞生长、细胞周期、细胞信号传导和细胞粘附过程中的作用。*Myc* 通过调节自身和其他基因的表达来影响正常细胞和肿瘤细胞的生长增殖，但其确切的机制不是很清楚。作者用含有 6 416 条基因和 ESTs 的芯片检测了 *c - myc* 激活在人主要纤维母细胞中所引起的基因表达变化，研究 *myc* 在细胞生长分化过程中的作用和功能。实验结果表明，27 个基因表达持续增加，9 个基因表达持续减少。这些基因表达的变化显示了 *myc* 可在多个方面影响细胞的生理过程，包括细胞生长、细胞周期、细胞粘附和细胞骨架的结构等。变化的基因中，包括有许多诸如真核起始因子 5A 一类的与细胞生长有关的核蛋白、核仁颗粒及纤维丝蛋白基因，提示 *myc* 可能是通过影响这些基因来调控细胞的生长。*c - myc*激活还影响到许多细胞周期基因的表达，从而造成细胞周期的改变，如 *myc* 上调了 G1 细胞周期蛋白 D2 和细胞周期依赖性酶结合蛋白 CKsHs 的基因表达，下调了细胞周期依赖性酶抑制因子 *P21*(*Cip*1)的基因表达。*Myc* 对细胞粘附和细胞结构的调节，是通过抑制编码细胞外基质蛋白纤维素、胶原和构成细胞骨架成分的肌球蛋白基因来实现的。对肿瘤坏死因子受体相关蛋白 TRAP1 的作用，可能是 *myc* 介导的细胞凋亡的机制。最后，两个免疫亲和素肽酰 - 脯氨酰 - 顺反式异构酶 F 和 FKBP52，也被 *myc* 介导上调，其中后者在细胞分裂中起着重要作用。另外，作者还建立了一种模式匹配(pattern - matching)方法来研究 *myc* 作用的靶基因，解读 *myc* 的功能，其基因表达谱与用基因芯片检测内源性的 *myc* 的基因表达谱相一致。

Schuldiner O 等比较了 *C - myc* 所致肿瘤和 *N - myc* 所致肿瘤的基因表达谱，并进一步分析在不同类型 *myc* 所致肿瘤中都有特异表达的基因，发现这些在 *myc* 所致肿瘤中过表达的基因大部分涉及细胞周期的调控。另外，通过分析这些基因调控序列中 *myc* 可能的结合位点，

发现其中的 *E2F1* 和 *TSC2* 基因可能为 *myc* 的靶基因。

Jo H 等用芯片研究了癌基因 *ras* 的功能。一些 *ras* 基因调控的信号转导途径，很大程度上严格控制了靶细胞的信号产生和表型，认识 *ras* 基因的功能，就必需要鉴别出 *ras* 基因作用的下游靶基因。芯片技术和差异显示技术极大地提高了鉴别 *ras* 靶基因的能力。

Rusiniak ME 等通过芯片研究维甲酸的靶基因表达水平。全反式维甲酸(ATRA)通过与融合癌蛋白急性早幼粒细胞白血病维甲酸受体 α(PML－RARalpha)结合，可导致急性早幼粒细胞白血病的发生，用芯片检测被维甲酸处理的、能表达 PML－RARalpha 的 TF1 髓样白血病细胞的基因表达变化，结果表明最初在内皮细胞中发现的，被肿瘤坏死因子 α 诱导表达的 *B94* 基因是维甲酸的靶基因之一。维甲酸处理 TF1 细胞 1 h 就可诱导 *B94* 基因表达，*B94* 基因也可在 NB4、UF1 和 HL－60 细胞中被维甲酸诱导表达，但在其他造血细胞中不被诱导表达。

Svaren J 等研究了 *EGR1* 反式作用因子在前列腺癌中的调控作用。作者将带有 *EGR1* 基因的腺病毒载体导入前列腺癌细胞系，然后用寡核苷酸点阵检测基因表达的变化。对可能是由 *EGR1* 调控的基因再通过定量 RT－ PCR 加以证实。在被鉴别出的 *EGR1* 靶基因中包括几种与神经内分泌相关的基因，如涉及前列腺癌发展的神经元特异性烯醇化酶基因等；此外，还发现几种生长因子也受 *EGR1* 调控，包括胰岛素样生长因子 II、血小板衍生生长因子 A、转化生长因子 β1 等。这种将芯片技术、腺病毒基因转染技术和定量 RT－PCR 技术相结合的思路，也为研究反式作用因子和其他肿瘤相关基因的靶基因及其功能提供了一个通用的策略。

Saramaki O 等通过荧光原位杂交和芯片技术，研究了编码真核翻译起始因子 3(eIF3) p40 亚单位的基因 *EIF3S3* 与前列腺癌发展的关系，结果表明晚期前列腺癌普遍存在 *EIF3S3* 基因的扩增，*EIF3S3* 基因在前列腺癌发展中发挥重要作用，预示前列腺癌的预后很差。

Calaluce R 等用芯片技术研究了层粘连蛋白－5(Laminin－5)在人前列腺癌中所起的作用。作者在其先前的研究中发现，细胞外基质和前列腺癌细胞之间的相互作用关系到前列腺癌的进展，层粘连蛋白－5 作为一种重要的基底膜糖蛋白，在前列腺癌中有缺失的现象，而层粘连蛋白－5 的受体－α6β4 整合素在前列腺癌中的表达也有变化。为深入探讨层粘连蛋白－5 在前列腺癌中功能和作用，用含 5 289 个基因的芯片检测了分别用纯化的层粘连蛋白－5 处理前列腺癌细胞系 DU145 各 0.5 h、6 h 和 24 h 后的基因表达变化。结果表明分别有 4、61 和 14 个基因在层粘连蛋白－5 处理 0.5 h、6 h 和 24 h 后发生变化，这些基因涉及细胞信号转导、细胞粘附、细胞周期和细胞结构等。实验结果经 Northern 杂交和免疫组织化学予以证实。

Wang L 等通过生物信息学和芯片技术在基因组范围内研究了 *P53* 基因的靶基因。首先，根据人类基因组计划的测序图和转录图，找出 4 852 条至少含有一个 *P53* 结合序列的基因，再随机挑选 13 个基因作 mRNA 分析，发现其中 11 个基因可受 *P53* 调控。通过染色质免疫沉淀印迹对 5 个基因的启动子分析，表明 *P53* 都可以在体内与它们结合。进一步用含 33 615 个基因的 cDNA 微阵列检测 *P53* 的靶基因，发现 1 501 个基因可对 *P53* 作出或上调或下调的反应。可见，用芯片技术可大大提高发现 *P53* 靶基因的能力。这种综合了生物信息学和全基因组范围基因表达分析的方法，为研究整个基因调控网络提供了一个新的手段。

Mayanil CS 等利用基因芯片发现了 *Pax3* 基因的下游靶标。*Pax3* 是一种转录因子，在胚胎神经管、神经脊和躯体衍化等发育过程中发挥着重要的作用。作者在先前的研究中发现，*Pax3* 在神经管母细胞瘤细胞系 DOAY 中过表达会导致 α－2,8－唾液酸转移酶基因(*STX*)表达上调，并使唾液酸等神经细胞粘附分子产生增多，提示 *STX* 可能为 *Pax3* 作用的下游靶基因。鉴于 *Pax3* 在个体发育过程中有多种重要的功能，有必要进一步明确 *Pax3* 的作用靶标及

功能。作者提取稳定转染了 *Pax*3 细胞的 RNA 与寡核苷酸点阵杂交，经与空载体对照和空白细胞对照比较后，发现总共有 270 个基因的表达在转染 *Pax*3 后会发生变化。寡核苷酸点阵数据再用独立的 cDNA 点阵和实时定量 PCR 检测加以确证，发现在一些有显著变化基因的启动子中含有成对的与 *Pax*3 类似的同源结构域。应用启动子－虫荧光素酶报告基因转染分析和电泳分析后，发现了包括 *STX* 在内的几种先前未报道的 *Pax*3 下游作用靶基因。

Richer JK 等研究了两种不同亚型黄体酮受体(PR)在乳腺癌细胞系中不同的生理功能。将两种不同亚型受体的基因导入乳腺癌细胞系，然后用芯片分析基因表达的变化。在 94 个黄体酮受体调控的基因中，有 4 个只由 PR－A 调控，有 65 个只由 PR－B 调控，只有 25 个是两种受体都可以调控的。临床上，这些受特异调控的基因可用于鉴别研究不同亚型的乳腺癌。

美国国立癌症研究所制定了肿瘤基因功能分析计划，采用包括基因芯片等在内的技术手段建立一个完整的人类 Unigene 系统，以供查询各类肿瘤相关基因的序列及其组织来源。该计划还从肺癌、乳腺癌、卵巢癌、前列腺癌和胃肠道癌等人类几种主要的癌组织中建立 cDNA 文库，以方便人们研究肿瘤相关基因的功能。

五、肿瘤诊断的研究

基因芯片也为肿瘤诊断提供了一个新的方向。通过检测体内大量基因的变化水平，特别是肿瘤相关基因的变化水平，或者检测肿瘤相关基因是否存在突变等异常，就可以诊断出人们是否患有肿瘤及肿瘤类型，或者判断病人已患肿瘤所处的阶段并预示其发展及预后，这也是基因芯片技术在临床实践上的又一个重要实际应用。

美国国家人类基因组研究室的 Hacia JG 等用高密度的寡核苷酸点阵，研究了与人类遗传性乳腺癌和卵巢癌密切相关的 *BRCA*1 基因的突变和多态性，被检测的 *BRCA*1 是其第 11 外显于，长达 3.45 kb。作者在芯片上固定了 96 600 个长度为 20 碱基的寡核苷酸探针，然后用标有不同荧光素的待检样本和对照样本与芯片共杂交，洗片后测荧光强度。在被检测的 15 例患者中，发现 14 例有点突变、插入突变或缺失突变，而 20 例对照样本均为阴性结果，另有 8 例单核苷酸多态性(SNPs)被检出。后来，Hacia JG 等经过多方改进，又设计制备了 96 600 个长度为 25 碱基的寡核苷酸点阵，并用来检测长度达 5.59 kb 的整个 *BRCA*1 基因(共含有 22 个外显子)的序列变化，取得了理想的实验结果。另外，Hacia JG 等还用高密度的寡核苷酸点阵，检测了许多其他基因的突变和 SNPs。

美国 Affimetrix 公司的 Cronin MT 等分别用两种 DNA 芯片检测囊性纤维化跨膜转导调节因子(*CFTR*)的突变，其中一个芯片含有 428 个探针，用于筛查 95 个碱基的 *CFTR* 基因第 11 外显子是否有变异或多态性，另一个含有 1 480 个探针，用于检测 *CFTR* 基因的已知突变，包括缺失、插入和碱基置换。最后，用芯片分析了 10 个未知病人样品的 *CFTR* 基因，其结果与 PCR 产物的限制性酶切片段分析结果完全一致。

Favis R 等将多重 PCR、连接酶检测反应(ligase detection reaction，LDR)和寡核苷酸微阵列结合起来，在一个反应管里可同时检测上百个突变，从而能够同时迅速分析大量样本。多重 PCR 是在一次 PCR 反应中加入多对引物，同时扩增一份 DNA 样品的不同序列，每对引物扩增的产物序列长短不同，根据长短不同的序列存在与否，就可检测是否有基因片段的缺失或突变。LDR 是一种用于分析基因序列中单个碱基突变的技术，通过一对紧邻引物与双链靶 DNA 中的一条单链退火，然后由 DNA 连接酶进行连接反应，就可扩增出大量单链 DNA 序列，若两个紧邻引物连接处碱基所对应的模板 DNA 碱基发生突变，会使引物末端不能与模板配对结合，则

引物之间就不能连接和扩增，结合 PCR 的 LDR 可通过一次扩增检测产物中的多个单突变位点。综合应用这几种方法，Favis R 等成功地检测出分类尚不明确结肠癌中 *K－Ras* 和 *P53* 基因中的突变，并且还很好地检出 *BRCA*1 和 *BRCA*2 基因中的低频突变。

人类恶性肿瘤中，约有 60% 与人类 *P53* 抑癌基因的突变有关，目前研究人员已经研制成功了可以检测 *P53* 基因所有编码区（外显子 2～外显子 11）错义突变和单碱基缺失突变的基因芯片。以外显子 7 的第 248 个密码子为例，野生型为 CGG，在芯片上做出 5 条探针，相应位点分别为 GAC、GCC、GGC、GTC 和 G－C，根据杂交后的荧光显色图，就可以分析出该位点为何种突变。

Lehman TA 等在检测家族遗传性乳腺癌中生殖系（germ－line）*P53* 基因内含子突变的频率及其功能作用时，采用了 Affimetrix 公司生产的 *P53* 基因突变诊断芯片进行研究。生殖系 *P53* 基因的突变在遗传性乳腺癌患者中所占的比例为 1% 或更少，作者用 42 例具有很显著乳腺癌家族史的乳腺癌患者为研究对象，研究了生殖系 *P53* 基因外显子和内含子发生碱基突变的频率及其功能的变化。42 例患者平均年龄为 44 岁（年龄跨度为 29～69 岁），其中 12 例（占 29%）为著名的 Ashkenazi 家系。采用 PCR－单链构象多态性（PCR－single strand conformation polymorphism，PCR－SSCP）分析技术、测序和基因芯片等方法，对生殖系 *P53* 基因第 2～11 外显子及其周围内含子的突变和多态性进行检测，结果表明在 3 例（占 7.1%）遗传性乳腺癌患者中，存在生殖系 *P53* 基因第 13 964 位的 G 变为 C（G13 964C）的突变，其中 2 例患者为杂合突变，1 例为纯合突变，而在 107 例非遗传性的乳腺癌患者中，无 1 例有 *P53* 基因的 13 964 GC 突变。42 例遗传性乳腺癌患者中，除 *P53* 基因第 6 内含子的 G13 964C 这一突变外，再无 1 例有 *P53* 基因第 2－11 外显子的突变，但谱系分析表明，3 例患者都持续伴有 Li－Fraumeni 综合征的多种肿瘤的家族史。对带有 *P53* 基因 G13 964C 突变患者的 *BRCA*1 和 *BRCA*2 基因进行检测，结果表明 *BRCA*1 和 *BRCA*2 基因没有伴随的有害突变，这种有害突变在其他遗传性乳腺癌患者中曾被发现过。用凋亡形态学发现方法和细胞生存检测方法，对 2 株来自有 *P53* 基因 G13 964C 突变患者的淋巴母细胞系进行功能检测，发现突变可延长顺铂处理后的细胞存活期并减低化疗介导的细胞凋亡。免疫组织化学检测到 *P53* 突变蛋白的高表达，证实 *P53* 基因 G13964C 的内含子突变为一个有功能的突变。

Wen WH 等用 *P53* 基因芯片和传统 DNA 序列分析法，分别检测 108 例卵巢癌的 *TP53* 基因突变时发现，总共有 77 例突变被两种方法之一检测出来，其中 71 例被芯片检测出来，63 例被传统 DNA 测序法检测出来。两种方法都检测出 57 例卵巢癌中相同的突变，另外检出 31 例卵巢癌中 *TP53* 基因为相同的野生型，一致率为 81%。不一致的结果中，有 14 例基因芯片检出而未被常规 DNA 序列分析法检出，有 6 例被常规 DNA 序列分析法检出而未被基因芯片检出。芯片检测法显示出 94% 准确性、92% 灵敏性和 100% 特异性，而传统 DNA 序列分析法显示出 87% 准确性、82% 灵敏性和 100% 特异性，基因芯片可用作一个检测 *TP53* 基因突变的有力工具。

目前，美国的 Affymetrix 公司开发的诊断芯片中，除检测 *P53* 抑癌基因突变的芯片外，还有能检测 HIV 逆转录酶基因、乳腺癌相关基因 *BRCA*1/*BRCA*2 和细胞色素－*P*450 基因等是否有突变的芯片，这些芯片都已经上市并且用于临床。

应用基因芯片技术进行肿瘤诊断时，通常首先要制备出针对各个肿瘤的诊断芯片，实际操作时，就是将具有诊断价值的各个肿瘤的相关基因集中起来点在一块或几块芯片上，再用这种有针对性的芯片与待检样品杂交，做到有的放矢，这样就可以大大提高突变的检出率和准确

率。国内已有一些肿瘤诊断芯片研制成功的报道，如复旦大学遗传所和联合基因公司等单位进行肺癌相关基因芯片的研制，就是首先有目的地搜集了 60 个肺癌相关基因，用修饰后的引物将其分别克隆，经纯化、变性后，用点样仪以微阵列的形式将其点布于醛基化的玻璃片上；然后将人支气管上皮恶性转化细胞模型 BEP2D 细胞的原代、20 代、35 代的总 RNA 逆转录后再经 LD PCR 标记成荧光探针与基因芯片进行杂交。荧光扫描仪扫描后显示，图像背景均匀、信号清晰，表明杂交效果良好，芯片制作成功。

将基因芯片技术与比较基因组杂交(CGH)技术相结合，还可高分辨率地检测到全基因组范围内 DNA 拷贝数的变化，为深入了解肿瘤的发生机制和进行肿瘤诊断提供技术支持。

CGH 技术是检测染色体或基因组 DNA 水平是否发生变异的有效方法，通常采用两种不同的荧光素标记 DNA 做探针，一种用红色荧光素(如 Cy5)标记肿瘤细胞 DNA，一种用绿色荧光素(如 Cy3)标记正常细胞 DNA ，两者混合后与正常人分裂中期的淋巴细胞染色体共杂交。根据比较两种探针所标记荧光信号强度的比率来判断肿瘤患者的 DNA 是否存在增加、缺失或复制等变化，如肿瘤 DNA 有增多，则杂交带显红色；如减少则显绿色；如没有变化则显黑色。但是 CGH 存在分辨率低、难以进行特异性基因定位及需要制备中期染色体等缺点，通常只能检测超过 20 Mb 染色体区域的拷贝数有无异常，使得该技术的应用受到一定局限。为克服这些缺点，Solinas - Toldo S 等将芯片技术与 CGH 技术相结合，推出了微阵列 CGH 技术，无需制备中期染色体，将荧光标记的基因组 DNA 与微阵列上的 DNA 序列杂交，可检测到 40 ~ 75 kb 大小 DNA 的数量变化。Pinkel D 等应用噬菌体人工染色体(PAC)和细菌人工染色体(BAC)克隆微阵列检测乳腺癌 20 号染色体缺失，分辨率达 40 kb 的缺失。Pollack JR 等采用 2 kb 的 PCR 产物制成 cDNA 微阵列，检测分析了乳腺癌细胞系和乳腺癌组织 DNA 拷贝数的变化，灵敏度和特异性达 98%，可高分辨率地鉴别全基因组范围内的基因扩增与缺失。Bruder CE 等则用这项技术快速而广泛地检测了大量 II 型神经纤维瘤患者中致病基因缺失突变的频率和范围。

Heiskanen M 等将 CGH 技术、cDNA 微阵列和组织微阵列结合起来，检测到成纤维细胞生长因子受体 2(fibroblast growth factor receptor 2 ，FGFR2)在乳腺癌中的低频扩增。乳腺癌细胞系 SUM - 52 中存在几个高水平 DNA 扩增位点，包括 *FGFR2* 所在的染色体 10q26 区域。首先，用 CGH 技术检测细菌人工染色体(BAC)克隆微阵列，发现 SUM - 52 中有 *FGFR2* 的高水平扩增。接着，用含 588 个基因的 cDNA 微阵列检测到 *FGFR2* 在 SUM - 52 中过表达超过 40 倍。最后，用荧光原位杂交(FISH)检测含 750 例原发性乳腺癌样本的组织芯片，结果表明 *FGFR2* 只在 1% 的原发性乳腺癌患者体内有扩增。

Hui AB 等应用基于芯片技术的 CGH 检测了多形性神经胶质母细胞瘤(glioblastoma multiformes ，GBM)中多个基因的扩增。对 14 例多形性神经胶质母细胞瘤样本(包括 7 例癌细胞株和 7 例癌组织样本)及正常对照样本，用切口平移法分别将其 DNA 标记上绿色或红色荧光素，再与用含 58 条癌基因的噬菌体人工染色体 P1、PAC 和细菌人工染色体 BAC 克隆制成微阵列芯片杂交，检测这 58 个基因的表达情况。结果表明有高水平扩增的癌基因包括 *CDK4*、*GLI*、*MYCN*、*MYC*、*MDM2*、和 *PDGFRA* 等基因，样本中扩增频率最高的癌基因包括 *PIK3CA* (64.3%)、*EGFR* (57.1%)、*CSE1L* (57.1%)、*NRAS* (50%)、*MYCN* (42.9%)、*FGR* (35.7%)、*ESR* (35.7%)、*PGY1* (35.7%)和 *D17S167* (35.7%)等基因。

Daigo Y 等应用简并寡核苷酸引物 - PCR(degenerate oligonucleotide primed - polymerase chain reaction, DOP - PCR)、激光捕捉微分离技术和点阵 CGH 技术检测了乳腺癌中多个癌基因的扩增，该技术可用于包括常规样本、甲醛固定样本和石蜡包埋样本等在内的多种样本形式肿瘤组

织中基因扩增的检测，并且经过与常规的中期染色体 CGH、点阵 CGH 和荧光原位杂交法相比较，证明该法能够可靠地检测基因的扩增。通过 28 例肿瘤样本与含 57 个癌基因的基因组芯片的杂交分析，作者从少量的样品就得到乳腺癌中癌基因的扩增谱，为深入了解乳腺癌发生的分子机制和临床表现提供了参考依据。

理想的肿瘤治疗方案应该依据病人准确的临床和组织病理诊断来制定，但一些肿瘤不典型的临床症状和组织病理形态特征又常常使之难以被确诊，基因芯片技术也可为这类问题的解决提供了一个可行方案。Ramaswamy S 等利用肿瘤基因的表达信息对多种肿瘤的诊断进行了研究，他们选择了 14 种常见肿瘤共 218 份样本和正常对照组织共 90 份样本与芯片杂交，对样本 16 063 条基因和表达序列标签(ESTs)的表达水平进行数据处理和分析统计后，总的诊断准确率达到 78%。

六、筛选抗癌药物和监测肿瘤治疗效果

随着人类基因组计划的完成，宣告了生物信息学时代的到来，大量的基因序列数据等待着人们去一一破解其中的奥秘。在抗癌药物和肿瘤治疗领域，我们同样面临诸多的生物信息学问题，如药物是否或如何影响肿瘤细胞的生长，肿瘤治疗效果及预后的评价等等，传统的研究手段对这些问题显得力不从心，这就促使人们寻找新的技术和方法来解决，正所谓工欲善其事必先利其器。

基因芯片技术以其高通量、微型化和自动化的特点，在抗癌药物的筛选、药物作用靶标的发现、药物的毒性评价、药物的时效关系及作用的分子机理等方面有着其他方法不可比拟的优越性。在很短的时间内，基因芯片就可筛选成百上千，甚至更多的化合物，从而大大提高新药的研发速度。另外，一种药物的作用有时是多方面的，基因芯片有助于发现一种药物的新的功能。如原先设想的某一药物是针对某一靶标的，但在全基因或广范围筛选中却发现该药在另一方面有很强的作用，从而将其开发成另一种新药。芯片技术可同时发现众多的新基因和新的靶分子用于新药的设计，大大促进制药工业的发展。肿瘤治疗药物 Herceptin 就是一种抗 *HER2/neu* 癌基因表达产物的重组人源化抗体，可用于 *HER2/neu* 癌基因过表达乳腺癌的治疗，并获得美国 FDA 的批准。

美国国立癌症研究所(NCI)的 Weinstein JN 等在其药物发现计划中，应用基因芯片技术筛查了 70 000 以上化合物对 60 个肿瘤细胞系的作用，对肿瘤有潜在治疗作用的化合物，进一步在 DNA、RNA、蛋白质和功能水平上研究其药理作用。通过基因芯片技术和生物信息学，他们发现了许多基因与基因、药物与基因间的相互作用，如在卵巢癌细胞中天冬酰胺合成酶与酶类药物 L－天冬酰胺酶就存在着相关作用，这些相互作用为药物的发现、药物设计以及肿瘤的个体治疗方面提供了借鉴。显而易见，用传统方法进行如此大规模的药物筛选是难以想象的，只有将高通量的生物信息处理技术与传统研究手段相结合才能使药物研究迈上一个新台阶。

Scherf 等用含 10 000 个基因的微阵列对 60 个细胞系的基因表达谱进行分析，获得一系列的标准曲线，发现大部分肿瘤细胞系基因表达特征与其原代肿瘤组织的表达特征相一致。应用生物信息学和化学信息学，进一步筛查 122 种对肿瘤细胞系基因表达有影响的化合物，从中遴选出真正有效的化合物作为抗肿瘤药物，然后再进行相应的毒理药理实验直至临床试验。

Clarke PA 等研究了肠癌患者化疗前和治疗期间肿瘤基因表达变化，发现丝裂霉素 C 和 5－氟尿嘧啶治疗均可使糖苷合成酶和尿嘧啶－DNA 糖基酶的基因表达增加。该研究提示，这类研究既有助于阐明药物的作用机制，也有助于确定药物作用的靶基因，为新药研究提供线

素。

Marton MJ 等利用基因芯片技术，快速检测了肿瘤患者对顺铂作用后的反应，发现肿瘤细胞受到药物作用 24 h 后 mRNA 表达发生改变，从而推测出该药物的药理作用及对不同肿瘤细胞的杀伤效应。

Kudoh K 等检测了阿霉素诱导型和阿霉素耐药型癌细胞的基因表达谱。作者将 MCF－7 肿瘤细胞被阿霉素短暂处理后的基因表达谱，与筛选出的 MCF－7 耐阿霉素型细胞的基因表达谱进行比较分析，结果发现被阿霉素短暂处理的细胞，有多种基因的表达随时间变化而变化，其中包括一些能被阿霉素诱导表达的基因，也在耐阿霉素型细胞中呈组成性高表达。这些实验数据可帮助人们阐明药物耐受的机制，并为寻找其他抗癌、抗肿瘤药物提供参考。

Chen CC 等用芯片检测了大豆异黄酮金雀异黄素对膀胱癌细胞的生长抑制作用。用 50mM 抑制剂量的金雀异黄素和不同长短的时间，分别处理膀胱癌细胞，提取处理和未处理膀胱癌细胞 mRNA 制成探针，与含 884 个已知序列的 cDNA 芯片杂交。结果差异表达的基因多数编码调控信号转导和细胞周期蛋白，这些基因可用于帮助研制抗肿瘤药物。

Zhou Y 等用 1 694 个肿瘤相关基因的芯片监测了拓扑异构酶 I 抑制剂喜树碱(CPT)对结肠癌细胞系 HCT116 基因表达的影响。为了使细胞反应一致，在 CPT 处理之前，先用抗病毒药阿菲迪霉素(APH)处理使细胞同步在 S 期，然后，分别用 20 nm 和 1 000nm 的 CPT 处理细胞，使细胞分别呈现出可逆的和不可逆的 G2 停滞期。将这两种剂量 CPT 处理后的细胞和未经处理的对照细胞与芯片进行杂交，结果两种剂量作用下的肿瘤细胞基因表达谱存在显著的差异。33 条基因在 CPT 处理的前 20 h 表现出特征性变化，取其中 5 条再用 Northern 杂交检测，实验结果与芯片结果比较后发现相当一致，平均相关系数为 0.86。几条由 *P53* 激活的应答基因在 1 000 nM CPT 处理或延长 APH 处理时间后上调，但上调基因似乎并不直接导致细胞周期停滞，因为通过延长 APH 处理时间导致的上调基因在去除 APH 后并不阻滞细胞周期。相反，与有丝分裂有关的一组基因在 CPT 处理后，其细胞周期依赖性的上调被延缓或抑制，这组基因表达上调的中断直接导致了 G2 期阻滞。另外，他们观察到一些下调基因在细胞周期被阻滞后恢复表达。这是一个应用基因芯片技术研究抗肿瘤药物对肿瘤细胞剂量－时效关系的例子，药物剂量的不同导致肿瘤相关基因的表达出现很大的差异，从中可推断出药物的剂量也是影响肿瘤治疗效果的因素。

Zembutsu H 等用全基因组 cDNA 微阵列，检测了 85 个人异种移植物对抗肿瘤药物的反应，以期建立一个能预测抗肿瘤药物对患者个体疗效作用或毒性作用的方法。将来自 9 种人体器官的 85 个异种移植物植入裸鼠，然后用 9 种抗肿瘤药物处理，这些药物包括 5－氟尿嘧啶、3－[(4－amino－2－methyl－5－pyrimidinyl)methyl]－1－(2－chloroethyl)－1－nitrosourea、氢氯化物、阿霉素、环磷酰胺、顺铂、丝裂霉素 C、氨甲喋呤、长春新碱和长春碱。85 个异种移植物对抗肿瘤药物的敏感性用含 23 040 个基因的芯片进行检测分析，结果表明 1 578 个基因在用药后变化，其中 333 个基因对 2 种或更多抗肿瘤药物有反应，32 个基因对 6 种或 7 种药物有反应。这些数据为肿瘤的个体治疗和发现或设计新的无耐受的抗肿瘤药提供重要的参考价值。

基因芯片技术也可用于肿瘤疗效及预后的评价，为肿瘤的用药、治疗方案的选择提供及时的参考。van 't Veer LJ 等应用基因表达谱预测乳腺癌的预后获得成功。乳腺癌是一种临床多发病，但是即使在同样的病情状态下，乳腺癌病人的治疗效果和预后常常会有很大的差异。对于乳腺癌组织是否转移扩散，采用诸如淋巴结和组织病理检查等手段有时很难作出准确判断。化疗和激素疗法虽然可以减低约 1/3 癌组织转移的可能性，但仍有 70% ~ 80% 接受治疗而存

活下来的病人可能会发生转移。Veer 等用基因芯片测得 117 例原发性乳腺癌的基因表达数据，通过分析处理这些数据，从中找到一批与乳腺癌预后相关的基因，包括与细胞周期调控、肿瘤浸润、肿瘤转移和血管生成有关的基因，用这些基因的表达水平作为乳腺癌预后的指标。临床实际检测结果表明，这种方法的预测效果超过了所有当前使用的临床指标，可为乳腺癌和其他肿瘤的治疗提供参考依据。

（黄　巍）

第五节　生物芯片在肝癌及病毒性肝炎诊断中的应用

一、基因芯片技术在肝癌研究中的应用

基因芯片技术也称基因微矩阵（microarray），是 20 世纪 90 年代发展起来的一种前沿生物技术，具有高通量、低消耗、自动化的特点。美国 Affymetrix 公司于 1991 年生产了世界上第一块寡核苷酸基因芯片，1995 年第一块以玻璃为载体的基因微矩阵芯片在美国 Stanford 大学诞生，这标志着基因芯片技术步入了广泛研究和应用的阶段。随着人类基因组计划（HGP）的完成，通过基因芯片技术能确定细胞内所有基因的表达谱，可同时获得成千上万个基因活化的模式。目前，基因芯片技术已被广泛地应用于临床医学各领域的研究，在肝病研究中也有一定的应用，并已取得了初步进展。基因芯片在肝病研究中主要应用于原发性肝癌、乙、丙型肝炎、肝硬化及肝纤维化等疾病的诊疗上。以下分疾病进行简单的介绍。

（一）基因芯片在原发性肝癌发病机理研究中的作用

基因芯片在原发性肝癌研究的应用方面有较多的报道。肝癌是一种常见的恶性肿瘤，其侵袭性强、预后差，居我国癌症死因的第 2 位。目前研究认为，肝癌的发生和发展是一个复杂的多阶段、多因素、多基因参与过程，实际是多种肝癌相关基因表达失常或多肿瘤抑制基因失活所致。正常基因的突变或缺失、癌基因的异常扩增和表达以及多个基因的协同作用、基因本身多效性和机体免疫因素，最终决定肝癌表型表达与否。DNA 芯片为研究肝癌发生发展中的基因开关及表达程度提供了强有力的工具，利用它可以随时获取肝癌细胞生长各期与肝癌生长相关基因的表达模式，因而基因芯片在肝癌研究领域中具有非常重要的价值。

1. 寻找肝癌相关基因。寻找和克隆新的肿瘤相关基因，以往采用的一个重要方法和途径是连锁分析和遗传分析技术。但是，对于像肝癌这类数量大遗传性不显著的多基因、多因素疾病的应用有一定的局限性，效果不理想。而基因芯片可以进行高通量的生物信息处理，确定细胞的基因表达谱，被检测目标 DNA 密度高，样本用量少，自动化程度高，便于大量筛出新基因，大大提高了发现新基因的效率。一些学者将基因芯片技术与差异表达技术有机地结合，通过对两种组织进行差异比较，提供差异表达基因的富集库，将这些克隆 cDNA 点到基因芯片上，可以同时、快速、重复地筛查上万种 DNA 分子，从而达到高效率、高通量、低成本地筛查和分析差异表达基因的目的，为研究在包括肝癌在内的各种肿瘤的发生、发展过程中的各个环节和阶段相互协调、共同表达的基因群提供了有力的工具，也使快速寻找和研究肿瘤相关新基因成为了可能。Eisen 等在进行大量的 DNA 芯片杂交后，对已知功能旧基因和未知功能的新基因的表达谱数据进行统计分析，建立基因聚类分析（gene clustering）的数学模型，可以确定不同基因在表达上的相关性，从而能够快速获得新基因的功能信息。

特异而敏感的肿瘤标志物是进行肿瘤的早期诊断、预后判断以及疗效评估的重要手段。

目前，原发性肝癌的诊断以甲胎蛋白(AFP)为主，但30%以上的乙肝病例AFP检查呈阴性。因此，寻找更特异、更敏感的肿瘤标志物是肝癌研究的一个重要课题。通过寻找和克隆新的肝癌相关基因，可望从中获取发现新的肝癌肿瘤标志物的线索。

2. 肝癌发病机制及基因表达谱研究。从过去单一因素致癌到多步骤、多因素综合致癌理论的形成，人们对肿瘤的发病机制的认识在最近几十年中取得了突飞猛进的发展。但是，由于技术条件的限制，以往人们研究肿瘤主要是针对几个基因，至多几十个基因，并试图以此来阐明肿瘤的发病机制，结果并不乐观。现在，伴随着人类基因组计划(HGP)和基因芯片技术等新技术的发展，我们可以大规模检测肿瘤组织内基因谱表达情况，将有可能真正阐明肝癌等肿瘤发生的分子机制。例如，*c - myc* 是正常细胞和肿瘤细胞中一个重要的癌基因，Coller等采用标有6 146个基因和表达序列标记(expressed sequence tags, ESTs)的寡核苷酸微阵列技术发现人纤维母细胞中 *c - myc* 基因的激活导致27个基因诱导表达、9个基因表达被抑制，这些基因与细胞生长、细胞周期、信号传导和细胞黏附等作用的调节有关。

肝脏在不同的环境影响下，有着不同的生长状态，这种状态的差异主要通过细胞内转录的RNA水平来表现。细胞内mRNA的转录水平反映了在不同的环境、细胞类型、细胞生长阶段和细胞状态下的功能信息，绘制所有基因的表达图谱对于研究基因的功能十分重要。利用cDNA表达谱芯片大规模、高通量和并行处理的优点，对细胞基因水平进行大规模平行研究，可以绘制出一张反映正常、异常和受控条件下所有基因表达的时空图，使人类的基因图从一维走向多维。并且，通过生物信息学方法，对各基因的表达谱进行比较和统计分析，确定不同基因在表达上的相关性，从而找到未知基因的功能和已知基因的未知功能。只有完全了解基因表达情况及其精细的调控才能解译细胞通路的动力学变化，包括表达瞬息的量变，不了解基因表达及其水平就不可能理解基因型和表型之间的相互关系。比如，感染不同肝炎病毒的肝癌呈现不同的基因表达形式，差异表现在编码代谢致癌剂或抗肿瘤药物的酶，同时判定了一定数量与恶性程度和浸润有关的基因，为选择新的治疗靶点、化疗方案提供线索。

Kawai等利用cDNA芯片进一步对产AFP肝癌细胞株的基因表达谱进行研究，结果表明产AFP肝癌细胞株具有许多种类特异的基因表达谱，这有助于理解其与不产AFP肝癌细胞株的因果关系，有助于研究基因群及其功能，为未来更合理的治疗提供依据。

HCV非结构蛋白(NSP)在病毒复制和肝病病理过程中起着重要的作用，cDNA表达谱芯片可观察NSP调节HCV对肝癌细胞基因表达的影响。用cDNA芯片检测HCV NSP诱导下肝癌细胞株的生长、代谢过程、细胞周期及转录因子的变化，发现这些表达的基因与病毒感染和肝癌形成有着密切的联系。

在国内，李瑶等人已利用基因表达谱芯片研究人正常肝和原发性肝细胞癌组织基因表达的差异性。他们将4 096条人cDNA点在特制玻片上制备成表达谱芯片；利用肝和肝癌组织的mRNA经逆转录方法将Cy3和Cy5两种荧光分别标记到两种组织的cDNA上，制备成cDNA探针，并与表达谱芯片进行杂交和扫描，共筛选出了差异表达的基因903条。所获数据将为鉴定参与肝癌发生的基因打下良好的基础。

另外，利用基因芯片表达谱数据还可随时获取肿瘤细胞生长各期与肿瘤生长相关基因的表达模式开始研制一种DNA芯片来获取肿瘤细胞生长相关基因的表达模式。该芯片的研制成功将为肝癌等疾病的发病机制及诊断研究作出重要贡献。

3. 肝癌相关基因功能研究。当前，基因研究的主要手段还是对基因进行逐个研究，即把一个癌基因或抑癌基因转到培养的细胞或转基因动物中，或将某一基因敲除(knock out)，观察

发生的生理或病理效应。但是，在生理或病理状态下，每一个基因都并非单独起作用的。人为地加入或剔除一种基因在很多时候并不能反映实际的情况。现在已经明确，参与基因表达调控、机体免疫及细胞分化等重要生命过程的许多相关基因是成簇排列的，它们可以分布在一较大的染色质区域或不同染色体上，在基因表达调控方面具有明显不同于单个基因的特性，其中的各个基因是作为统一的整体进行调节，各基因在细胞分化、个体发育过程中保持高度的时空以及表达量的协调性。基因芯片具有高通量、低消耗、自动化的优点，它将成千上万个基因固定在一块芯片上，可以对来源于不同个体、不同细胞周期、不同分化阶段、不同病变、不同刺激的肿瘤组织或细胞内的 mRNA 进行检测，迅速将某些基因与肿瘤联系起来，缩短了实验周期，大大加快了这些基因功能的确定。

Sgroi 等利用基因芯片，结合应用激光捕获显微切割技术(laser capture microdissection)，检测在同一名乳腺癌症患者的在体纯化正常细胞、浸润性癌细胞以及转移性癌细胞的基因表达水平，围绕乳腺癌的发生和发展过程，研究处于不同病理过程的癌细胞中各种基因功能的情况，得到了较为可信的结果。在肝癌发生、发展的复杂过程中，很可能同样存在着这样相互协调的共表达的基因群体。将肝癌生长的各个环节和阶段中共同表达的众多基因作为一个“基因群”(gene cluster)进行研究，筛选在肝癌相关的多个环节上起作用的分子和起关键性调控作用的基因，揭示肝癌在基因水平的本质，将对于控制肝癌发展、肝癌的有效治疗有着重要意义。

(二)基因芯片在原发性肝癌诊断和治疗研究中的作用

1. 肝癌分子病理分型和最佳治疗方案的选择。目前肝癌的分类主要依靠于临床资料和组织形态学标准，但是这种方法存在明显的不足。在临床上常出现一些组织病理学表现相似，而临床转归和治疗效果有显著差异的肝癌病例。这些肝癌病例的各个亚型，由于在临床上表现十分相似，单纯依靠临床症状往往难于区分，而传统的突变基因的筛查又十分耗时耗力，而且不同亚型之间对于相同的治疗方案的反应十分不同，所以正确地分析亚型，对于肝癌的治疗以及预后情况，具有十分重要的作用。尽管传统肿瘤的分类方法已经沿用了近 30 年，但目前尚无一种能够辨别新的肿瘤类型的通用方法。基因芯片技术的出现，使基于广泛的基因表达分析的肿瘤分类方法成为可能。这种对肿瘤分类方法的研究可以分成二类，一类是发掘肿瘤类型，即确定以前未被认识的肿瘤亚型；另一类是预测肿瘤类型，即将特定的肿瘤标本归于已定义的可以反映当前状况和预后的肿瘤类型中。

对于肿瘤分型，目前已经有了一些肿瘤分型基因芯片。通过在芯片上固定同一种肿瘤不同类型的特征性基因，我们可以得到某种肿瘤的分型基因芯片。另外，除了开发新的基因分型芯片外，基因表达谱芯片也可以用于肿瘤分型。Gloub 等运用基因芯片技术建立了一种基于基因表达类型的肿瘤分类方法，并已试验性地应用于人类急性白血病。作者用发掘肿瘤类型的方法自动地找出了急性类细胞性白血病(AML)和急性淋巴细胞性白血病(ALL)之间的区别，而预测肿瘤类型的方法能自动地确定新的白血病类型。结果证明了这种完全基于检测基因表达类型的肿瘤分类方法是切实可行的，并暗示了这种研究方法可以不依赖于以往的生物学知识，发掘其他肿瘤的新的亚型和预测肿瘤类型的研究。这种发掘肿瘤类型的技术能够用于辨别所有肿瘤的主要亚型，这种“亚型”与普通的病理分型相比，具有真实和可重复性的优点。而且这种分子病理分型方法能够用于研究一种涵盖不同类型肿瘤的基本分类机理，更重要的是，预测肿瘤类型的技术今后有望应用于临床，提供一种客观的通用技术预测肿瘤患者的生存期和预后。因此，在不久的将来将这一技术应用于肝癌研究，将对发掘出新的肝癌亚型，或者更准确地预测出肝癌的类型，从而预测出肝癌患者的生存和预后以及对指导肝癌患者的用药具有重

要的意义。

Shirota 等应用双色荧光标记判断肝癌的差异表达基因，发现肝癌组织中 22 条差异表达基因与肝癌分化程度有关，其中 13 条基因为与细胞分化和生长相关的转录生长因子或组织特异性蛋白基因，如 *DADI*、*IGFBP* - 2、*CyclinG*2、*E*2*f*3 等。这些基于临床样品检测的信息，不仅有利于进一步阐明肝癌的发生机制，而且有助于确定肝癌分类、指导治疗和判断预后。

Stanford 大学 Brown 和 Stanford 教授进行了一项实验研究，从几百例肿瘤标本中获取了数以万计的基因的转录本。Dudoit 等人则在这项研究的基础上开发出了一种多元统计学方法对基因芯片的表达谱进行分析。这种统计方法能够对包括肝癌在内的多种肿瘤的分子差异进行分型，统计分析的最终目的是希望能在基因表达谱数据的基础上发展出一种新的肿瘤的分类法或"分类学"(taxonomy)。统计分析包括一系列群集分析和判别式分析。

另外，Liversey 等则应用基因芯片检测单个细胞基因表达谱。他们使用 PCR 扩增技术对大鼠视网膜上单个特定神经元中的总 RNA 进行扩增，并合成和扩增 cDNA，应用含 130 个已知基因和 960 个从成年大鼠视网膜 cDNA 文库中随机挑选克隆的基因芯片进行杂交。每一种类型细胞的芯片杂交数据均以同一份全脑 cDNA 标本作为参照物，并对细胞类型进行聚类分析。为了直接筛选出细胞类型特异性基因，作者还对每一种细胞类型的芯片数据进行了两两比较，辨别那些仅在各自不同类型细胞中表达的基因。结果证实了这种在单个细胞水平上进行基因表达研究的可行性，这也对包括肝癌在内的肿瘤分子病理分型研究有借鉴意义。

2. 肝癌早期诊断芯片。我国的肝癌患者以及每年的新发病例均居世界首位，利用这些丰富的肝癌疾病资源，分离、定位和克隆一批在我国发生率高，危害性大的肝癌致病基因和相关基因，进而研究其在肝癌早期诊断和治疗方面的应用，将具有重大的社会意义。基因芯片可以用于检测基因突变。从正常人基因组中分离出 DNA 并与 DNA 芯片进行杂交，获得标准图谱。通过对两种图谱的比较、分析，就可以得出病变的 DNA 信息。它不仅可以准确地确定突变位点和突变类型，更主要的是它的快速高效是目前直接测序法所无法比拟的。

Hacia 等已经利用 DNA 芯片检测遗传性乳腺癌和卵巢癌基因 *BRCA*1 第 11 个外显子的基因突变，在 15 例患者样品中，发现 14 例有基因突变，突变方式包括点突变、插入及缺失等。在 20 例对照样品中均未发现假阳性。使用该芯片可一次检测出 *BRCA*1 基因的所有突变位点。1997 年，Ficarrahe 和 Eversole 等合作研制 DNA 芯片检测和肿瘤相关的 *P*53 基因全长编码序列，一般认为 50% 以上的肿瘤患者该基因会发生突变。美国 Affymetrix 公司宣布新的 DNA 芯片能同时检测 250 种和癌症相关的基因。因此，不论肿瘤的病理变化涉及的是单基因或多基因病变，都可以从病理与健康细胞、组织基因的差异表达中发现致病原因的线索。可以预见，利用肝癌诊断芯片检测正常人群中肝癌相关基因的差异表达，以达到早期诊断、早期治疗肝癌的目的的技术不久将得以实现。这项技术在肝癌及其他肿瘤的诊疗中应用前景是很广阔的。

3. 肝癌耐药基因的筛选、新药的开发以及药物抗肝癌作用机理的检测。对于肝癌病人，尤其是晚期患者来说，目前尚未找到根治的办法，所以选择合适的药物来控制病情的发展至关重要。但是，长时间地使用同一种药物，常会使肝癌细胞对于药物产生耐药性，影响治疗效果。这样，就要求我们一方面鉴定分离出肝癌耐药基因，同时还要明确抗肝癌药物的作用机制、靶位点，并据此开发出更有效地治疗肝癌的新药物。利用基因芯片技术，我们可以找出肝癌特异性相关的致病性基因，并且通过研究相应的抑制药物治疗后引起的基因表达变化，观察肝癌细胞或患者用药之后的反应，从而为开发抗肝癌的新药开发开辟一条新路。

Marton 利用基因芯片技术，能快速检测出肿瘤患者对顺铂作用后发生的反应，发现肿瘤细

胞受到药物作用后 24 h 后的 mRNA 表达发生改变，从而推测出该药物的药理作用及对不同肿瘤细胞的杀伤效应。郑世荣等则利用基因芯片技术研究血卟啉单甲醚（HMME）光动力学疗法对人原发性肝癌 HEPC2 细胞基因表达的影响，以阐明其作用的分子机制。观察结果显示，光照射后实验组细胞呈现凋亡形态；芯片扫描显示 2.47% 的检测基因（389/1 538）出现明显表达差异，其中 80% 的基因表达下调，这些基因多与增殖调控功能有关；而细胞凋亡通路中几个重要关键蛋白（CCP32、AIF、Mch2）的基因明显上调。研究表明 HMME 激光疗法诱导细胞凋亡的机制可能是线粒体调控。美国国立癌症研究所在癌症治疗发展计划中观察了 65 000 种以上化合物对 9 种肿瘤组织来源的 60 个肿瘤细胞系的作用效应，Scherf 在所得数据的基础上利用 1 000 个基因芯片评估药物对 60 个肿瘤细胞系的基因表达的影响，从中筛选出抗肿瘤药物候选化合物，并对其做临床药效评价。在不久的将来，利用基因芯片技术筛选出针对性更强、更有效的抗肝癌药物是完全可能的，而且与传统的基因筛选方法相比效率将大为提高。

二、基因芯片技术在病毒性肝炎及其他肝病诊断中的应用

（一）病毒性肝炎的流行病学

迄今为止已发现数十种病毒可引起肝炎性损害，其中较为明确的有甲、乙、丙、丁、戊、己、庚七种。我国是病毒性肝炎的高发区，有约二亿肝炎病毒携带者，每年感染人数呈上升趋势。其中以引起慢性肝炎的乙型、丙型肝炎病毒对患者的危害最大，是发展成肝硬化、肝功能衰竭和原发性肝癌的主要危险因素之一。

乙型肝炎是由乙肝病毒（HBV）引起的、以肝脏炎性病变为主并可引起多器官损害的一种传染病。本病广泛流行于世界各国，一般说来，乙肝表面抗原（HBsAg）的携带率热带地区高于温带，男性高于女性，儿童高于成人，城市高于农村；且各地区的人群感染率有很大不同。北美、西欧及澳大利亚为低度流行区，人群中 HBsAg 的阳性率仅 2% ~ 5.0%，乙肝核心抗体（抗 - HBc）的阳性率为 4% ~ 6%，儿童与新生儿感染者较少见；东欧、日本、地中海地区、西南亚及独联体为中度流行区，人群中 HBsAg 及抗 - HBc 的阳性率分别为 2% ~ 7% 和 20% ~ 55%，儿童和新生儿的感染较为常见；我国和东南亚国家及热带非洲为高度流行区，HBsAg 和抗 - HBc 的阳性率分别为 8% ~ 20% 和 70% ~ 95%，儿童与新生儿感染极为常见。根据 1990 年的流行病学调查资料表明，我国人口中约半数感染过乙肝病毒或正在感染中，有 10% 以上的孕妇是乙肝病毒携带者，如不接受乙肝疫苗预防，则 60% 的新生儿在两年内可被感染上乙型肝炎。乙肝亚型的分布随地域不同亦各不相同：*ayw* 多见于非洲、东地中海、中东及巴基斯坦；*adw* 多见于北欧及美洲；*adr* 多见于东南亚及远东，*adw* 及 *adr* 多见于印尼、马来西亚、新几内亚及泰国。而我国则具有一定的民族性，汉族人以 *adr* 为主，*adw* 次之；蒙古族、维吾尔族、回族等兄弟民族几乎都是 *ayw* 亚型，*ayr* 亚型极其少见，而 *adr* 亚型几乎没有。

丙型肝炎是由丙型肝炎病毒（HCV）引起肠道外传播的一类传染病。输血及血制品引起传播是本病的主要途径。据国外报道，输血后引起的肝炎中，90% 以上是丙型肝炎；10% ~ 25% 散发性急性病毒性肝炎为丙型肝炎。我国由于缺乏系统的流行病学资料，故确切发病人数尚不清楚，但输血后丙肝的比例约占 60% ~ 80%。而急性散发性肝炎患者中，丙肝约占 12% ~ 24%。此外，因单采血浆回输血球发生血液污染而引起丙肝暴发流行的报告亦较多。近年来，靠输血途径传播的丙肝例数有逐渐增多的趋势。如美国 1988 年的调查数字表明，急性丙肝中 46% 有静脉注射毒品史，10% 有不正当性接触或家庭内肝炎接触史，5% 有经常接触血液或血液透析史，而输血引起者仅占 6%。其中不正当性行为与家庭内肝炎接触史的发病率要比不

具备这些因素者分别高出 11 倍和 6 倍。此外，母婴垂直传播也是一个途径。丙肝的这种多途径传播，造成了预防上的困难，故其发病率正呈逐渐上升趋势，流行范围也逐渐扩大。此外，由于人群中存在较多的慢性 HCV 携带者也可通过各种途径传播 HCV，从而有可能造成丙型肝炎的发生和流行。

(二)现有病毒性肝炎的诊断方法

目前对病毒性肝炎的诊断主要是对甲、乙、丙、丁、戊等肝炎病毒的筛查上，其手段是传统的血清学诊断及基因诊断。

由于甲型肝炎和戊型肝炎无慢性化，且有较好的血清学诊断指标，故一般不需进行基因诊断。乙型肝炎病毒抗原，特别是表面抗原，在血液中的浓度较高，通过抗原检测即能明确诊断。但在对患者进行抗病毒治疗时，由于血清中病毒核酸含量是反映病毒数量和复制活跃程度最可靠的直接指标，因此，病毒含量的检测就显得必不可少，它是动态观察药物疗效的确切指标，特别是对于 HBeAg 阴性患者。对于丙型肝炎病毒感染的诊断，现行的第三代血清学检测试剂虽有较好的敏感性和特异性，但约有 20% 患者可不出现抗体，或有的患者抗体存在时间可以很长，因此，对其进行病毒核酸的检测以明确是否存在现行感染，是非常必要的。HCV 感染后，由于抗体出现较晚，要早期诊断也只有进行核酸的检测。因此，核酸检测在 HCV 感染的诊断中要比 HBV 感染的诊断重要得多。至于 HGV 和 TTV，因至今未能明确其对人类的致病性，基因诊断在临床上也较少应用，目前主要用在基础和实验研究中。

最初临床上应用最多的 HBV DNA 检测方法为斑点杂交，它有较好的稳定性和特异性，也不需要昂贵的设备，但它是一种定性检查，且敏感性较低(检测敏感度约在 10^5 copies/mL 以上)，在一定范围内不能反映出病毒含量的变化，随着 PCR 检测技术的发展日趋成熟，临床应用斑点杂交检测法越来越少。PCR 方法敏感性高，可检测出极微量的核酸，因操作技术等方面的影响，其实际敏感性一般约几百个 copies/mL 左右。由于影响 PCR 测定的因素颇多，常易出现假阳性和假阴性结果，因此，必须根据临床上的其他资料，进行综合分析。由于斑点杂交和 PCR 定性均不能反映病毒含量的动态变化，难以对抗病毒药物的疗效作出确切的判断。20 世纪 90 年代初，由 Chiron 公司推出了第一代支链 DNA(bDNA)检测试剂，其线性范围为 $10^5 \sim 10^9$ Eq/mL。bDNA 检测的敏感性在 HBV 为 1 000 copies/mL 以上，而在 HCV 检测其最低限度为 4.8×10^4 copies/mL。虽然它能反映血液中病毒含量的变化，但试剂价格十分昂贵。在 HBV DNA 定量检测中，线性范围的下限较高，HCV RNA 的定量检测中敏感性不够，难于在临床上广泛应用。90 年代后期，Digene 公司推出了第二代杂交捕获法检测 HBV DNA 的试剂，通过化学发光捕获法和信号放大技术检测 HBV DNA 含量，敏感范围在 $5 \times 10^3 \sim 5 \times 10^9$ copies/mL，且检测时间短，批内和批间变异小，特异性高，应当说较适用于临床的 HBV DNA 定量检测，但目前尚未推广应用。现今发展成熟的实时定量 PCR 方法，具有省时、费用相对较低，而且能反映血液中病毒核酸的确切数量等优点，目前也正在临床上广泛应用。但不同的检测系统的敏感性，即检测的最低病毒核酸量仍有差异，同样也存在因测定操作而出现的误差，有时误差还是较大的，此时往往要观察病毒核酸变化的趋势，以对临床结果做出正确判断。

此外，有关基因芯片、生物芯片技术用在病毒性肝炎诊断，已初露头角，它是将大量特定序列的核酸片段或蛋白有序地固定在载体上，与标记好的待检核酸或蛋白分子进行反应，通过检测荧光信号的强弱而判断样本中的靶分子数量，从而实现对化合物、核酸、蛋白质、细胞及其他生物组分的准确、快速和大信息量的筛检。它具有高度平行性、多样性、微型化和自动化的特性，以往常用的任何方法都难以与之相比。目前常用诊断芯片主要是基因芯片和蛋白质芯片

两类。

(三)生物芯片技术在病毒性肝炎诊断中的应用

1. 用于病毒免疫标志物的平行检测。在引起肝炎的诸多病毒中,许多不仅传染途径相似,而且还可重叠感染。譬如,乙型、丙型肝炎病毒都可通过体液、血液及血液制品传播,而且可以重叠感染,重叠感染的患者较单一感染的患者发生重症肝炎和死亡的比率要高。

目前常用免疫学方法来检测肝炎病毒的标志物,该方法的局限性在于抗原抗体的检测是建立在单向反应的单一指标上,其能获得的信息量十分有限,如需获得大量的疾病信息势必要进行多次试验。如果利用蛋白质芯片则要方便快捷得多。

蛋白质芯片技术原理类似于常规的酶联免疫反应原理,即将特异性抗原或抗体固定于载体,待测样本按比例稀释后与其上的抗原或抗体进行反应,再加入荧光标记的抗原或抗体,用计算机软件对荧光信号进行分析,即可获得准确的定性或定量结果,一张芯片上可分布上千甚至数万个抗原或抗体,并能标记多种荧光素,使之能同时被快速检测,且可进行病毒的血清学分型。我国学者吴凌凌将乙肝表面抗原和 e 抗原分别作不同的稀释,点于一张 2 cm × 3 cm 的 PVDF 膜上制成芯片,用于同时检测表面抗体和 e 抗体,结果表明两组抗原和抗体反应相互间无影响,此外还可用于滴度测定。整个检测过程仅需 2.5 h,如果二抗用常用的辣根过氧化物酶来标记,肉眼就可观察结果,方法简便、特异、灵敏。进一步有可能制成集成型检测芯片。深圳益生堂公司研究开发的丙型肝炎病毒抗体蛋白芯片检测试剂盒是国内目前惟一获得中国药品生物制品检定所出具检定合格报告的产品,也是国际上第一个获药品权威机构认可的生物临床诊断用芯片。它具有操作简单、信息量大、并行检测,高效省时的特点。

2. 用于肝炎病毒的分类、分型、变异和定量检测。为遴选抗病毒药物,评价临床疗效,除生化、免疫和病理指标外,还应明确病毒的核酸结构和体液病毒核酸的含量,用 DNA 测序、PCR 等法虽可达到目的,但技术难度及实验成本限制了其应用。基因芯片则可对所有的肝炎病毒的分型、变异、突变和病毒核酸含量进行高通量、平行检测。

利用基因芯片技术可以同时检测多种肝炎病毒,例如上海博华公司研制开发的乙型、丙型肝炎双检基因芯片,是把 HBV 和 HCV 高度保守的各长约 200 bp 的基因片段作为靶基因,点于特制的玻璃片基上,同时点上用于系统监测的与 HBV 和 HCV 靶基因没有同源性的植物基因片段作为阴性内参照基因,在血清抽提和荧光标记时分别加入检测系统的两段植物基因片段分别作为 HCV 和 HBV 的阳性内参照基因。从病人血清抽提的核酸经过 RT - PCR,将 PCR 产物标记上 Cy5 即可与芯片进行杂交,杂交信号由扫描仪扫描,再经计算机分析即可得出诊断信息。其最大的特点是能对乙肝和丙肝同时诊断,提高了检测效率,降低了检测成本,这对 HBV、HCV 混合感染及丙型肝炎急性期的早期确诊有重要的意义。最近,河南省研制了一种能同时检测乙型、丙型、庚型三类肝炎的基因诊断芯片,更值得可喜的是它们还开发了一种能临床应用、能定性定量分析的基因芯片分析系统。此外,武汉大学也正在研制一种便携式病毒性肝炎集成诊断芯片,它是在玻璃基片上分布有分成七个不同区域的点阵,在每个点阵,依次分别固定有甲、乙、丙、丁、戊、己、庚七种类型的病毒性肝炎的具有诊断意义的特征基因,通过表面分子杂交技术和电化学信号转换技术,从而可准确地同时检测七种病毒性肝炎。该芯片一次集成操作可同时完成七种病毒性肝炎的定性、定量检测,方便快速地确定有几种病毒感染以及各自的感染程度如何等,适用于医学诊断。

芯片技术还可以进行病毒基因分型的检测,如乙肝病毒就有多种基因类型,不同的基因类型之间存在致病性上的差异,同时,病毒的基因变异还可引起抗药性的不同。因此,临床医生

不仅要知道病人是否患有乙肝，而且还要确定其属于乙肝的哪种基因类型和突变类型。最近，由南开大学和南开基因工程有限公司利用基因研究成果及 DNA 芯片技术开发研制一种乙肝病毒基因诊断芯片，记录着 21 个质量控制信息和 429 个肝炎病毒基因类型，只需抽取一滴血，就可利用 400 多个探针对乙肝病毒基因的不同区域、多个位点同时进行检测，检测的信息量大，准确率可达到 100%，检测不但细微到病毒的基因亚型而且还可随时跟踪基因突变产生的结果，为医生用药提供信息，目前这项技术在世界上处于领先水平。

Livache 等人将经吡咯修饰的 5′ - pyrrole - TTTTTCCAAGAAAGGACCCG - 3′；5′ - pyrrole - TTTTTCTCCAGGCATTGAGC - 3′；5′ - pyrrole - TTTTTCAACCCAACGCTACT - 3′三个片段经电化学方法固定于氧化硅片基上的 48 个微金电极上，制成用于丙型肝炎病毒基因分型的诊断芯片，抽提待检样品 RNA，经过生物素修饰的特异引物的 PCR 反应即可与芯片杂交，整个杂交反应只需 15 min(45 ℃)。该芯片具有较高的检测灵敏度与特异性，无交叉污染。

3.用于病毒的耐药性突变检测（见本章第六节）。

4.结合核酸类抗体配基筛选技术(SELEX)，实现基因芯片技术对肝炎免疫应答中特异性标志物及相关蛋白质的分析。随着 DNA 结合蛋白研究的深入，受组合化学、抗体库和随机噬菌体肽库技术的启发，人们构建了随机核酸库并从中筛选出与靶蛋白特异结合的核酸配基，此技术称为 SELEX。与蛋白类抗体相比，此配基具有许多优越性：作用的靶分子范围更广，配基与靶分子的结合能力与特异性更强，动力学参数可依体外诊断条件的要求而改变，不受抗原毒性和免疫原性的限制，特异性和亲和力不受组织或样本中非靶抗原的干扰，体外人工合成实现标准化生产，合成时可随意连接其他功能基团和分子等。目前已从核酸库中筛选出各种与蛋白、核酸、小分子肽、氨基酸、有机物、金属离子等特异性结合配基，并应用于临床治疗和诊断。凡涉及抗体的诊断领域，几乎均可用核酸配基替代。选择针对肝炎免疫应答过程中特异性标志物及相关反应蛋白筛选核酸配基，人工合成后结合于固相载体，制作基因芯片，即可实现芯片技术的病毒免疫标记物及免疫应答过程中的细胞因子、细胞周期、细胞凋亡等免疫学检测。基因芯片技术虽有诸多优点，但要成为实验室或临床可以普遍采用的技术目前尚有一些关键问题亟待解决，如如何提高芯片的特异性，简化样本制备和标记操作程序，增加信号检测的灵敏度和高度集成化样本的制备，基因扩增，核酸标记及检测仪器的研制和开发等，已成为当今国内外研究的热点。

(四)基因芯片在肝纤维化和肝硬化研究中的作用

基因芯片技术在肝纤维化及肝硬化的诊治方面同样已初步发挥了作用。Shackle 等先后应用基因芯片研究原发性胆汁性肝硬化(primary biliary cirrhosis，PBC)和丙型肝炎后肝硬化，发现许多差异表达基因，这些基因涉及炎症、纤维化、细胞增殖、信号传导和凋亡等诸多方面。结果显示，PBC 组织中 Th1 和 Th2 型免疫反应分子皆表达上调，纤维化相关基因以接触生长因子和 $TGF\beta_3$ 为特征。另外，Wnt 和 notch 硬化组织中，表达上调的免疫反应分子以 Th1 型为主，*EMMPRIN*(*CD*147)、*CD*167 以及 Wnt 通道基因分泌型凋亡相关基因 3 为新发现的肝硬化基因。此外，运用 cDNA 微矩阵检测了正常人肝、纤维化肝和活化的 HSC 基因转录的变化，分析发现，上调最显著的是已知在纤维化中起重要作用的 I 型胶原、α—SMA 和键生蛋白。

以上研究有力地推动了肝硬化、肝纤维化诊疗技术的发展。相信在不久的将来，将有更多的基因信息借助基因芯片被大量获取。

以 Affymetrix 公司为代表的基因芯片技术发展迅速，将生物学、计算机科学、数学、化学、物理等各门学科融合在一起，开创了广阔的应用领域。作为一种强有力的探索未知表达基因的

新方法，该技术被应用于肝病研究，而且在短短的几年时间内已取得了较大的进展，但是许多方面仍有待完善、深入。例如，许多芯片实验以细胞株或动物为研究对象，其获取的生命信息应用于检测临床标本的特异性、敏感性及应用价值的评价需要一个过程。而以人体标本为研究对象，其结果对于解释人类疾病更直接、更具有实际意义，但组织标本是多细胞的混合体，这会对判定特定细胞中的基因表达情况有所干扰。幸运的是，细胞精确分离技术的发展日新月异，如激光捕捉微切割技术的出现，为有效解决这一问题提供了强有力的支持。

目前，通过基因芯片发现的肿瘤异常表达相关基因，与既往发现的癌基因和抑癌基因类似，仍缺乏特异性，寻找组织特异性肿瘤标志物的工作尚待继续；隐藏在已发现和已知的这些肿瘤相关基因背后，还可能存在着一个或多个决定集体发生肿瘤与否的关键因素，这些尚未知晓的因素急需深入研究。另外，我们不能停留在单纯寻找疾病未知新基因的水平，还应该进一步对差异表达 cDNA 的功能进行研究，对其生物学作用进行验证，最终协助肿瘤的临床诊断和治疗。

随着人类基因组草图的初步完成，生命科学研究已经跨入"蛋白质组学"的新时代，蛋白质的研究将成为生命科学领域继人类基因组计划之后的最重要、最前沿的战略课题。当此之时，蛋白质芯片研究发展更为迫切。由于基因水平不能直接反映疾病状态，最终起作用的仍然是蛋白质水平的异常表达。所以，结合蛋白质芯片技术研究肝病更有实际价值，尤其是在寻找肿瘤生物学标志物方面具有特别的功能。目前，这一研究领域的发展是十分迅速的。

（林菊生　张德太）

第六节　基因芯片技术在微生物变异及疾病耐药性检测中的应用

一、生物芯片在微生物基因变异检测中的应用

生物芯片及其技术是融微电子学、生物学、物理学、化学、计算机科学为一体的高度交叉的新科技，一旦被广泛应用于临床，它将在一些感染性疾病、重症传染病、恶性肿瘤、自身免疫性疾病、神经性疾病等方面，发生巨大的革命性的变革，基因技术在疾病的诊断、检测及治疗用药等方面有广阔的应用、研究价值。

生物芯片在微生物感染性疾病和重症传染病等疾病的临床诊断方面具有独特的优势。与传统检测方法相比，它可以在一张芯片上同时对多个病人进行多种病原的检测，无需机体免疫应答反应期，能及早诊断，待测样品用量小，能特异性地检测病原微生物的亚型及变异。

（一）微生物变异机制

1. 细菌变异的机制。细菌基因组序列出现变异是它的遗传基本特征之一，它是细菌进化和适应环境的必然结果。这些变异有可能导致细菌某些生物学特性的改变，从而改变其致病性和对药物的敏感性。

（1）突变。细菌以无性二分裂方式进行繁殖，理论上讲 DNA 复制过程十分精确，子代与亲代的基因组应是完全相同的。但在少数情况下，子代细胞会出现可以通过复制而遗传的 DNA 结构的改变，称为突变（mutation）。突变可分为点突变和多点突变，前者只有一个碱基对的变化，后者有两个以上碱基对的变化。点突变可以是碱基置换、碱基插入或碱基缺失。细菌的点突变往往是碱基置换，包括转换（transition）和颠换（transversion）。碱基的插入或缺失影响三联体密码的阅读框架，引起移码突变。多点突变时往往涉及广泛的染色体重排、倒位、重复或缺

失。

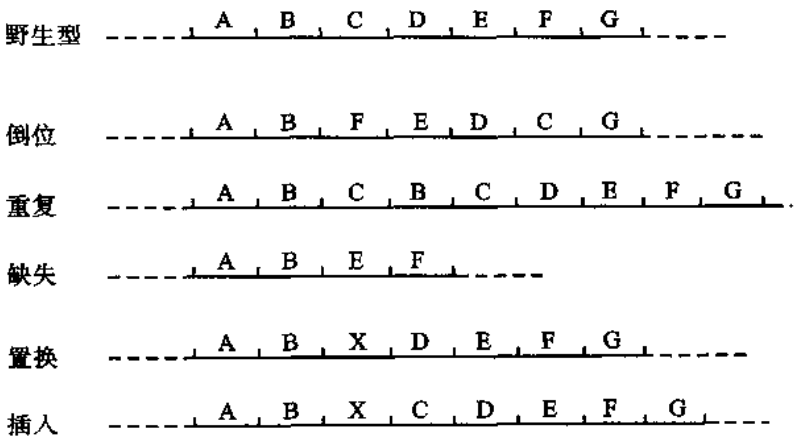


图 17-22 突变类型示意图

细菌群体中能够存活下来的一些突变株，可以被特定的环境条件所选择。进入病人体内的少量细菌经过生长繁殖也会自发地产生各种突变株，这些突变可以赋予细菌耐药性、增强细菌毒力或抗原性的改变，从而提高了细菌在病人体内的生存能力，致使突变菌株迅速过渡生长而被选择出来成为优势型。

(2)基因转移与重组。细菌基因转移和重组的方式有转化、接合、转导、溶原性转换和原生质体融合。

1)转化：受体菌直接摄取供体菌 DNA，从而获得新的遗传性状(如耐药性)的过程称为转化(transformation)。

2)接合：细菌通过性菌毛相互连接沟通，将质粒或染色体的 DNA 从供体菌转移给受体菌的过程称为接合(conjugation)。

耐药性质粒的接合传递极为常见，以 R100 研究得较为详细。R100 质粒由两部分组成，一是耐药性传递因子(resistance transfer factor, RTF)，能编码性菌毛，使其以接合方式传递；另一是耐药决定子(resistance determinant)，赋予宿主菌耐药性。一个耐药决定子可携带多个耐药基因，因此，携带耐药性质粒的细菌可同时对多种抗菌药物耐药，即多重耐药性。接合性耐药性质粒(R 质粒)通过接合方式可以在同一种属细菌间或不同菌属间进行传递，这在革兰阴性菌中更为突出，是细菌耐药性迅速播散，耐药株不断增加。

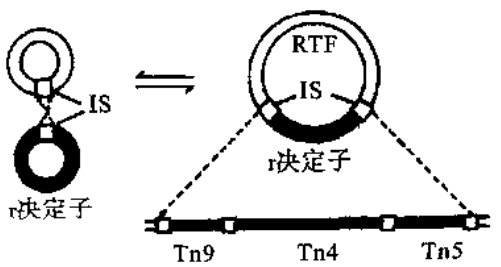


图 17-23 R 质粒结构图

细菌耐药性的遗传物质基础：细菌耐药性(drug resistance)是指细菌不能被抗菌药物常用剂量达到的血药浓度所抑制，因而临床治疗无效。耐药性的程度一般以某种药物对细菌的最低抑菌浓度(minimum inhibitory concentration，MIC)表示。从遗传学角度，细菌耐药性分为固有耐药性和获得耐药性。

固有耐药性(intrinsic resistance)：指细菌对某种药物的天然不敏感，也称为天然耐药性。这种耐药性始终如一，是细菌的种属特异性决定的。例如，两性霉素 B 能够与真菌细胞膜的固醇类结合，改变其通透性，而起到抗真菌的作用。细菌细胞膜缺乏固醇类，故对两性霉素 B 属于固有耐药性。同样，革兰阴性菌具有外膜屏障，决定了这类细菌对多种药物不敏感。因而固有耐药性是可以推测可知的。

获得耐药性(acquired resistance)：是由于细菌 DNA 的改变而出现了耐药性的表现，包括染色体突变和质粒介导的耐药性。

染色体突变：所有细菌的群体经常发生自发的随机突变，只是频率很低，其中有些突变赋予细菌耐药性。突变的频率与抗菌药物的使用无关，但药物形成的选择性压力有利于耐药突变株的存活，最终成为优势群体。例如，链霉素的作用靶位是细菌核糖体 30s 亚基上的 P12 蛋白，当细菌染色体上的 *str* 基因突变后，P12 蛋白的构型发生改变，使药物不能与其结合故产生细菌对链霉素的耐药性。

质粒介导的耐药性：1959 年日本学者 Akiba 和 Ochiai 分别做实验，将具有多重耐药性的大肠埃希菌与敏感的志贺菌混合培养，发现多重耐药性可由大肠埃希菌传递给志贺菌，首次证明了 R 质粒的接合传递。此后，在世界各地报道了耐药性质粒的普遍存在和广泛传播。耐药性质粒广泛存在于 G^+ 和 G^- 细菌中，几乎所有致病菌均有耐药性质粒。它们在菌细胞之间可以通过接合、转导和转化的方式进行传递，环境中抗菌药物形成的选择压力有利于耐药性质粒的播散和耐药菌株的存活。R 质粒在肠道菌中更为常见，推测其演变过程可能是耐药性传递因子(RTF)与耐药性基因或非接合性耐药性质粒结合形成多重耐药的接合性质粒，转座子的参与加速了 R 质粒的演变过程。多重耐药菌株所致的感染给临床治疗带来极大的困难。

3)转导：转导(transduction)是以噬菌体为媒介，将供体菌 DNA 片断转移到受体菌内，使受体菌获得新的遗传性状。根据转导 DNA 片断的范围，可分为普遍性转导和局限性转导。

4)溶原性转换：溶原性细菌因染色体上整合有前噬菌体而获得新的遗传性状称为溶原性转换(lysogenic conversion)。

5)原生质体融合(protoplast fusion)：是分别将两种细菌经处理失去细胞壁悬于高渗培养基中保持原生质体状态，然后将两种细菌的原生质体混合，滴加聚乙二醇促使原生质体融合。融合后的双倍体细胞可以短期生存，在此期间染色体之间可以发生基因的交换和重组，获得多种不同表型的重组融合体。

2.病毒变异的机制。

(1)突变：病毒基因组核苷酸序列发生改变，而不是由于他种病毒或生物体遗传物质的介入所引起的变异称突变。病毒增殖过程中常发生自发突变。因为 DNA 病毒与真核和原核细胞中 DNA 复制遵循同样的校读错误及更正的机制，故自发点突变率大致相同，约为 10^{-8} ~ 10^{-11}。由于 RNA 病毒复制中无校读机制，故 RNA 病毒自发点突变率远高得多。

突变按核酸的改变分类，最普遍的是点突变、缺失性突变、插入突变。突变也常按表型变化分类，如温度敏感突变、冷适应突变、耐药突变等。突变的表型表达不仅取决于突变的种类和位置，还取决于其基因的活性，即突变表型表达不仅可由于同一位置上碱基回复突变而被抑

制，而且还可被同一或不同基因上其位置出现的抑制性突变所抑制。

(2)重组：当两个有亲缘关系而性状不同的病毒同时感染同一细胞时，可以出现基因互换并产生兼有两个亲本特性的子代，称为重组。在新合成的核酸中会出现几种基因重组类型，即分子内重组、重配、复活。

1)分子内重组(intromolecular)：常发生于不分节段基因组病毒间，不同但有亲缘关系的两种病毒核酸分子发生断裂并交叉连接，核酸分子内部序列重新排列，如脊髓灰质炎病毒。在少数情况下，分子内重组亦可出现在两个无亲缘关系的病毒间，如 SV40 与腺病毒。当持续感染 SV40 恒河猴细胞在感染腺病毒时，由于 SV40 DNA 和腺病毒 DNA 同时整合入细胞 DNA 中，不仅出现互补作用即 SV40 作为流产性感染的腺病毒的辅助病毒，还可出现两病毒基因分子内重组的杂合子基因，此基因组被包裹于腺病毒核衣壳中。许多 DNA 肿瘤病毒在与宿主细胞 DNA 重组过程中，病毒携带的原癌基因会将被感染细胞转化为致瘤状态。通过分子内重组，逆转录病毒原病毒 DNA 于宿主细胞 DNA 整合是其复制循环的必要过程。

2)重配(reassortment)：常发生在具有分节段基因病毒间的基因重组，是由不同亲代病毒产生的子代核酸各个分子间的交换。因这些病毒的每个节段相当于一个基因，能独立复制，最后才组合纳入衣壳中。因此，当两株病毒在同一细胞内复制时，他们的核酸节段就可能随机分配而构成重组子代病毒，这种重组的发生频率远高于不分阶段病毒。

3)复活(reactivation)：同种大量灭活病毒间能经基因重组而出现感染性病毒子代称多重复活作用(multiple reactivation)。例如以紫外线灭活的病毒，则可能发生多重复活作用，因此，紫外线照射不宜于制备灭活病毒疫苗。交叉复活是指一个感染性病毒和灭活的亲缘病毒之间的基因重组，而不是不同的病毒原型与这一病毒的 DNA 片断间的重组。

(3)基因产物间的相互作用。

1)互补作用(complementation)：当两株病毒感染同一细胞时，由其中一个病毒株提供另一个病毒增殖所必需而不能生成的基因产物，因而使后者能增殖。病毒的遗传型不会改变。子代病毒具有亲本的基因型，虽然可能有一些非亲本的结构蛋白，如另一个病毒特异性的外壳。

2)表型混合(phenotypic mixing)：一种病毒体的衣壳或胞膜可包在另一种病毒体的核酸外面，称为表型混合，因为不是遗传物质变异，所以是不稳定的。例如几种肠道病毒感染细胞时，常发生衣壳转移，甚至有由两亲代编码的壳粒混合所组成的衣壳。有包膜病毒之间发生的表型混合导致产生相嵌包膜，例如甲型、乙型流感病毒感染细胞所产生的子代，可来自两亲代的刺突。因病毒衣壳与胞膜都可发生表型混合，若在自然界也发生这种现象，这可能改变病毒的宿主范围，并使某些病毒感染的血清学诊断发生困难。

3)多倍体(polyploid)：除逆转录病毒为二倍体外，所有脊椎动物病毒均为单倍体。但在借助胞膜芽生成熟的病毒中如副粘病毒，如果不同株病毒同时感染同细胞，许多基因组的子代病毒是异源性多倍体，他们也带有表型混合的胞膜抗原。

(二)微生物基因变异检测

1.微生物基因突变检测技术。生物种系为适应日益多变的生存环境，必须要进行自我的“调节”和“修饰”来维持自身种系的延续，微生物也不例外。因此发现并鉴定这些“修饰”了的病原微生物对临床正确诊断和治疗疾病是刻不容缓的事情。

微生物的鉴定经历了直接镜检与培养、生化反应、免疫学与血清学诊断和基因检测(PCR)的发展历程。基因检测是 20 世纪 70 年代末期伴随着基因重组技术而发展起来的一项以分析核酸中的碱基序列特征为基础的新的鉴定策略，即通过对某一微生物的某个特定的基因和

(或)其转录物进行分析和检测,以辅助临床医生对某种疾病进行诊断。与传统的鉴定方法相比,基因鉴定的优势在于更直接、更准确、更灵敏地反映出微生物的致病性,因此能以更高的灵敏度对疾病进行早期诊断,有利于疾病的及时防治。随着人类基因组计划的发展,越来越多的病原微生物基因组全序列被阐明,特别是其与生物芯片的结合,将大大加快病原体基因分型、鉴定和耐药性检测的准确性和效率,并使病原体的基因分类、鉴定和临床诊断等工作更全面、更可靠,避免了传统方法选择基因片段上的片面性,对监控再现和新现病原体提供了重要的技术手段。基因突变检测的核心技术是 DNA 分子杂交和聚合酶链式反应,近年来在这两种核心技术的基础上又衍生出许多新的检测手段,尤其是 DNA 芯片的出现使突变检测变得规模化和程序化。以下对近年来常用的几种突变检测技术加以描述。

(1)检测技术分类:

1)根据目的基因是否被放大分类:可分为杂交法和扩增法。前者被检测的目的基因不被放大,如分支链 DNA 技术,但用杂交法可以进一步检测经过放大了的基因片段,如 PCR 产物杂交分析;后者包括 PCR 及连接酶链反应,Qβ 复制酶系统等。

2)根据被测基因的核酸类型分类:① 在杂交法中,Southern 印迹用于检测 DNA;Northern 印迹用于检测 RNA;斑点印迹或称斑点杂交,既可检测 DNA 也可检测 RNA。② PCR 法主要用来直接扩增 DNA,但是通过逆转录技术,可以将 RNA 先转变为 cDNA 再行扩增,称逆转录 PCR,如检测 HCV、HGV、HIV 等。③ 原位杂交,或原位 PCR,也可分 DNA 或 RNA 原位杂交和原位 PCR 或原位逆转录 PCR。

3)根据是否提高了敏感度分类:分支链 DNA 技术(b - DNA),就是根据固相杂交的原理,采用了一种放大标记探针,目的探针不被放大,但检测信号被放大,从而提高了敏感度;套式或巢式 PCR,是用两套引物(内、外引物)作两轮 PCR,第二轮 PCR 是以第一轮 PCR 的产物作为模板,敏感度进一步提高,常用于极微量病毒基因检测,如检测 TT 病毒,以及检测大多数 RNA 病毒(称逆转录套式 PCR)。

4)根据诊断目的或要求分类:分为定性诊断、定量诊断和半定量诊断,以及序列分析等。

(2)方法和原理:

1)杂交法:

原理:杂交技术的基本原理是相同的。病原体基因的一级结构中都含有四种碱基:DNA 为腺嘌呤(A)、胞嘧啶(C)、鸟嘌呤(G)和胸腺嘧啶(T),RNA 为腺嘌呤(A)、胞嘧啶(C)、鸟嘌呤(G)和尿嘧啶(U)。在双链 DNA 或 RNA 中这四种碱基以固定关系配对,即 A - T(或 A - U),C - G,这种碱基配对关系就构成了 DNA 或 RNA 两条链的互补关系,也就是说,两条链之间是严格地通过 A 对 T(或 A 对 U)、C 对 G 的关系结合在一起的。在一定的温度下,这两条链可以打开,变成分离的两条链,但又可以在一定温度下结合起来,这种"再结合"仍然遵循互补原理,但是再结合时的两条互补链是随机碰撞而形成双链的,也就是说再结合后的两条链已经不再是"原配"的了,"杂交"一词就由此而来。如果我们在一条链上"打上标记",并用它去与另一条链杂交,如果另一条链的确存在,就必然会发生杂交反应,杂交体(新的双链)上也就会被同样带上了标记,再用特定的技术可以探测到标记信号。通过这个原理,我们就可以用已知的突变的耐药基因单链作为探针与样本中病原体的未知突变基因杂交,以了解病原体的耐药性突变,指导临床用药。

方法:杂交法中最关键的技术是制备标记探针。标记探针的制备过程包括探针的制备和探针的标记两部分,可分别亦可同步进行。常用的方法有切口平移法、随机引物法、末端标记

法、单链探针制备和标记等。目前还有应用比较广泛的寡核苷酸标记探针，即在寡核苷酸化学合成过程中将标记物(放射性的或非放射性的)掺入或修饰在探针上。① 斑点杂交：将待测标本用点状加样法，直接滴加膜上，然后用碱溶液使核酸变性并结合在膜上，再经中和及烘烤(尼龙膜可用紫外线照射)后，与探针进行杂交。② Southern 印迹：又称凝胶电泳印迹转移杂交。是电泳技术与杂交技术结合的一种方法。先将待测标本病原体 DNA 分离，经限制性内切酶消化成一系列片段，进行琼脂糖凝胶电泳，各片段因分子量不同而彼此分开，然后经碱处理凝胶，使 DNA 的片段变性，在原位将单链核酸吸印到硝酸纤维膜上，经烘干、固定，以放射性核素标记 DNA 探针进行杂交，最后洗膜和放射自显影，从显影区带鉴定待测标本 DNA 片段。③ Northern 印迹：是指 RNA(主要 mRNA)的分子杂交技术。其原理和操作过程基本上与 Southern 印迹法相同。差别在于电泳标本是 RNA 片段，使用的探针是标记的 DNA。④ 原位杂交：是细胞学技术与核酸杂交技术结合的一种特殊技术。用标记核酸探针，与细胞涂片、压片、细胞悬液滴片或组织切片上具有互补序列的单链 DNA 或 RNA 形成双链结构，然后通过放射自显影或酶底物显色或激发荧光等方法检测特定的核酸分子所在的部位。由于本法不需从细胞中提取核酸并可直接观察待测核酸所在的细胞类型、细胞内分布状态与细胞染色体的关系，因此常是检测病毒与细胞关系的极好方法。⑤ 微反应板杂交：杂交反应的载体不是硝酸纤维膜或尼龙膜，而是微孔反应板，操作方便，敏感度高。此法快速、简便，不需先抽提核酸，一次可检测大量标本，已广泛用于检测血清或肝组织中 HBV DNA、HCV RNA、HIV RNA 等。该技术近年来发展较快，预计将取代斑点印迹法。

2)扩增法：PCR 法是最常用也是最成熟的方法。

原理：利用耐热的 DNA 聚合酶(如 *Taq* DNA 聚合酶等)在体外条件下，催化一对引物间的特异 DNA 片段合成。它包括变性、退火、延伸三个过程。这 3 个过程组成一个循环周期，上一个周期合成的产物又可作为下一个周期的模板，如此循环往复，经过 n 个循环后，靶 DNA 的拷贝数呈 2^n 增长，可把靶 DNA 的拷贝数扩大百万倍以上。

方法：① 标本处理：作为 PCR 的待测标本，单链、双链 DNA 或 RNA 均可。如以 RNA 为起始标本，则需先通过逆转录获得 cDNA 第一链后才能扩增，此即逆转录 PCR(RT－PCR)法。标本中不能混有任何外源和内源性蛋白酶(可降解 *Taq* DNA 聚合酶)、核酸酶、DNA 多聚酶的抑制剂或能结合 DNA 的蛋白质。② 引物设计：由于引物决定扩增片段的特异性，因此它必须对致病微生物是特异的。可通过计算机进行致病微生物基因资料分析，先选出致病微生物基因组中具有相对独立性与保守性的序列，再设计与该序列两端互补的寡核苷酸片段进行人工合成。③ 扩增条件：须注意 dNTP 浓度、*Taq* 酶的质量、引物用量、镁离子浓度、反应温度与时间、实验环境、器材质量等，这些都影响 PCR 结果的质与量。尤为重要的是要严格预防产物污染，以避免假阳性。④ PCR 产物检测：根据检测目的不同可以选择琼脂糖凝胶电泳染色法、Southern 印迹转移、酶切图谱分析法、序列分析法和 PCR－ELISA。

3)基因定量诊断：基因定量是在基因定性诊断技术基础上发展起来的，但历史不长，应用范围亦有限。目前主要用于数种病毒基因的定量分析，如乙型肝炎病毒、丙型肝炎病毒、艾滋病毒(HIV)等。基因定量诊断可用于病情评估、疗效预测和观察及预后判断等，故有重要的临床应用价值。随着基因定量技术和方法的不断完善，其应用范围也将不断扩展。

杂交定量分析：① 分支链 DNA 信号放大技术：该技术系美国 Chiron 公司的专利。检测系统中的主要成分是 5 组功能不一的人工合成的寡脱氧核苷酸探针，其中标记放大探针是技术的关键。b－DNA 技术有以下优点：灵敏度高。其 cut－off 值为 700 000 Eq/mL；阳性检出率高。

Chiron 的第二代 b－DNA 系统将检测探针的针对性扩展到 HBV 基因的各种亚型和突变型；稳定性和可重复性好；操作简便、省时、易于普及。国内已有一些医院采用 b－DNA 技术进行 HBV 基因和 HCV 基因检测。② 液相分子杂交：分别由美国的 Abbott 实验室和芬兰的 Orion 公司建立。其基本原理是将放射性同位素掺入标记探针，与待检基因在液相中杂交，然后经过吸附洗脱或过柱洗脱，分离游离和结合探针，测定结合探针部分的放射活性。③ 双抗体夹心杂交法：采用两种抗 DNA/RNA 杂交体的抗体：一抗为捕获抗体，直接吸附在微孔反应板上，二抗是标记抗体，偶联有 AP，可通过酶促化学发光反应检测 DNA/RNA 杂交体的信号。④ 微反应板酶联杂交：方法与 b－DNA 相似，但在信号放大探针制备方式上有其独到之处，获得了国家发明专利。目前主要由于定量分析 HBV 基因组，敏感度比 b－DNA 略低，但不需要特殊仪器，可普及化等是其优势。与上海绿谷伟业生态工程公司合作生产的试剂盒不久将在国内问世。

PCR 基因定量技术：基本原理同 PCR 定性技术，方法上计有终点稀释法、外参照法、荧光 PCR、PCR－ELISA 等。竞争 PCR 的基本原理则是在定性 PCR 的基础上，在扩增反应管内加入了一种称之为内参照的竞争模板，可以与野生模板即目的片段等效地被扩增。内参照是决定分析系统是否可靠和稳定的关键因素。

2.生物芯片在微生物基因突变检测中的应用。面对众多的基因突变，传统的 PCR 和 DNA 测序方法费时费力，且成本较高，很难适应快速、经济、大规模及程序化的序列比较和突变分析的需要。

自从 1989 年首次报道了寡核苷酸阵列在检测已知基因组 DNA 序列变异中的应用以来的 10 多年时间，各国科学家把这一崭新技术扩展到生物领域的各个方面，如临床医学、药学、动物学、植物学、微生物学等等。如果已知突变的基因序列，就可在 DNA 芯片的表面，以微阵列的方式固定寡核苷酸（携带了突变位点）或 cDNA，使用这种 DNA 芯片就可以对研究者感兴趣的突变基因或突变基因表达进行测定。高密度寡核苷酸阵列的出现开辟了一条检测更复杂系统的途径。具体方法如下：

（1）4 L 砌砖方法。杂交测序法曾经在芯片出现的初期认为有可能取代 Sanger 创建的核酸测序法，但在经过实践后却否定了这一尝试。因杂交接合的条件与序列中碱基组成有关，难于取得一致的结果。对于易发生突变的已知核酸序列区段采用 4 L 砌砖方法进行再测序（resequencing），可以成功地鉴别完全错配和单个碱基错配，检测出突变位点，其方法是：根据已知序列片段，依次（每个向下移一个碱基）合成一定长度小片段寡核苷酸探针（L），每个小片段寡核苷酸中同一位置的一个碱基是可变的（A、G、T、C），即每一个寡核苷酸有四种不同的序列，以不同的寡核苷酸探针作为列，四种可变的碱基为行，有序地点阵于基质上制备成芯片。当经过标记的样品与芯片杂交后可以容易地判读突变的位点和序列。其方法如图：

在芯片上分为 1～6 列，每列均合成或点加序列为 1～6 的 15 mers 的寡核苷酸探针。在每一列中探针序列只有中间 1 个有下划线的核苷酸是不同的，它的碱基载于左侧。当这样的芯片与带荧光的样品杂交时，可以在芯片中出现序列特异的杂交结果（黑色方块）。荧光测定结果很明显地显示出野生型（序列为 TACGCA）变异型（序列为 TACTCA）。只要有 1 个碱基的突变，引起的错配可以使与点突变相邻的变异序列杂交信号明显减弱。

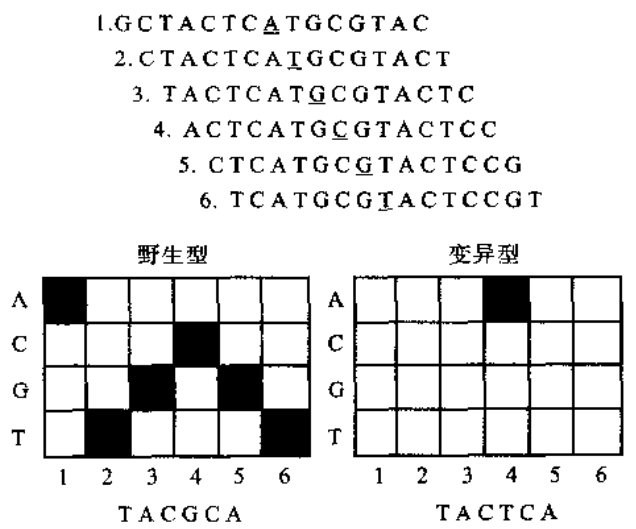


图 17－24　砌砖方式用寡核苷酸阵列芯片作 DNA 再测序鉴别点突变

(2)杂交测序技术(sequencing by hybridization,SBH)。SBH 是一种创新性的 DNA 测序方法,是建立在特定序列的 DNA 与特定长度的寡核苷酸集合杂交的基础上,未知 DNA 序列可以通过鉴定与靶 DNA 序列形成完整的双链体寡核苷酸的重叠区域来测定。

对耐药性微生物的突变体和多态性的检测,SBH 技术可大规模地检测和分析 DNA 的变异及多态性。Guo 等人利用结合在玻璃支持物上的等位基因特异性寡核苷酸(ASOs)微阵列建立了简单快速的基因多态性分析方法。将 ASOs 共价固定于玻璃载片上,采用 PCR 扩增基因组 DNA,其一条引物用荧光素标记,另一条引物用生物素标记,分离两条互补的 DNA 链,将荧光素标记 DNA 链与微阵列杂交,通过荧光扫描检测杂交模式 SHB 技术,即可测定 PCR 产物存在的多种多态性,该方法对人的酪氨酸酶基因第 4 个外显子内含有的 5 个单碱基突变进行分析,结果显示单碱基错配与完全匹配的杂交模式非常易于区别。这种方法可快速、定量地获得基因信息。α－地中海贫血中变异的检测也论证了该方法的有效性和可信性。Lipshutz 等人采用含 18 495 个寡核苷酸探针的微阵列,对 HIV－1 基因组反转录酶基因(*rt*)及蛋白酶基因(*pro*)的高度多态性进行了筛选。微阵列中内部探针与靶序列的错配具有明显的不稳定性,据此可快速区别核酸靶的差异。例如检测序列 5′GTATCAGCATXGCCATCGTGC 中 X 碱基的种类,可用下列 4 种探针 3′ ACTCGTAACCGTAGC, 3′ AGTCCTACCGGTAGC 3′ AGTCGTAGCGGTACC, 3′ AGTCCTATCGGTAGC。高密度探针阵列可检测具有特征性的较长序列相关的多态性与变异。筛选1 000个核苷酸序列的变异与多态性需要 4 000 个探针。用 100 μM 合成位点,设计1.28cm^{2}阵列,将有大约 16 000 个探针,能筛选 4 kb 序列。HIV－1 基因组中 *rt* 与 *pro* 在疾病过程中易于发生多种变异,这些变异将导致病毒对多种抗病毒药物包括 AZT、ddI、ddC 等出现明显抗性,因此检测和分析 *rt* 与 *pro* 的变异与多态性具有重要的临床意义。

(3)其他方式。

1)杂交信号增益法(gain of hybridization signal approach):信号增加的途径,允许所有可能的序列多样性的 DNA 片段进行部分扫描。用来判断长度为 N 的目标 DNA 双链的所有可能的单核苷酸碱基置换,需要设计包含有 8^N个探针,探针的长度通常是 20～25 个核苷酸。在单链的特定位置出现的四个可能核苷酸互补的 4^N个探针(分别称为“A”,“C”,“G”和“T” 探针)。探针

的设计通常在靶标的中心部位的碱基，提供了杂交特异性最佳的识别位点。为了确定靶标双链所有的碱基缺失，需要有 2^N个探针。但是，要确定长度为 N 的双链靶标中所有可能的 X 个碱基的插入，则需要 $2^N(4x)$个探针。如果要筛选长度为 10kb 的双链 DNA 靶标中所有可能的单核苷酸置换，需要 80 000 个探针；筛选所有可能的 1～5 个碱基缺失还需要再加 100 000 个探针；而筛选所有 1～5 个碱基的插入需要 27 280 000 个探针。用分析普遍的联合文库的芯片来判断整个序列的变化可以缩短探针的长度，减少探针的数量，但付出的代价是提高探针的特异性和敏感性。虽然，目前在芯片的生产技术上有所改进，但搜索一个以上碱基插入通常没有实际的意义。

原理是采用叠瓦式探针设计，将与突变位点上各种可能突变的序列互补的寡核苷酸探针固定在芯片上，以 PCR 扩增并标记靶序列，然后与芯片杂交，根据杂交信号的有无（即芯片上哪些突变探针上产生了强的荧光信号），可确定突变类型。这些探针获得的相对信号增强表明某个序列发生了改变。

```
5'... TGAACTGTATCCGACAT... 3'
      TGACATAGGCTGTAG
      TGACATCGGCTGTAG
      TGACATGGGCTGTAG
      TGACATTGGCTGTAG
       GACATAAGCTGTAGA
       GACATACGCTGTAGA
       GACATAGGCTGTAGA
       GACATATGCTGTAGA
```

图 17－25　叠瓦式阵列探针设计

297 bp 的整个 HIV－1 蛋白酶基因编码序列，应用该方法得到的杂交信号来筛选 167 个病毒单体所有单个核苷酸的变化。荧光标记的 RNA 碱基与包含有 12 224 核苷酸的两张芯片杂交。与单核苷酸置换的探针杂交的基本信号赋予核苷酸特性。分析 114 个样品 33 858 bp 的 HIV－1 序列，结果显示 98.26% 的数据集有一致性。从生物物理的观点来看，在基因组 DNA 中筛选杂合子的变化，双脱氧测序和芯片测序有同等的功效。

2）杂交信号减弱法（loss－of－signal approach）：在信号减少的途径中，评价序列多样性是依赖于参试样本和野生型的参比靶标与完全匹配的寡核苷酸探针之间杂交信号削弱的相对量。从理想化的状态来说，纯合子的序列变化导致了与完全匹配探针之间的杂交信号会大幅度削弱，从而判断在某一片段附近序列发生了变化。由于杂合子序列的多样性，与杂交野生的靶标杂交信号的强度回减低 50%，理论上将发现完全匹配的探针信号确定旁边的序列发生了变化。

其原理是设计与基因突变区完全配对的野生型寡核苷酸探针（一般为 25 nt）并固化在芯片上，扩增待测样品和野生型参照样品的靶序列并以不同的荧光进行标记后混合在一起与芯片杂交，比较两种杂交信号的差异（即分析待测样品的杂交信号相对于参照样品信号减少的相对量），可确定突变发生的大致区段，类似于比较基因组杂交。特点是可同时检测 dsDNA 的两条链，任何形式的突变信号都可检出。但该方法的一个缺点是不能确定突变类型，必须通过对杂交信号减弱区进行双脱氧测序来鉴定序列变异的类型。其敏感性、特异性优于信号增益法。

削弱信号的分析适合用于对任何序列的多样性进行实际筛选。用这一方法，设计芯片确定两条目标链 N 个 bp 所有可能的序列变化，最少包含 2N 的重叠的探针。例如，包含有 11 000 条寡核苷酸的芯片是筛选 5.5 kb 的 *BRCA*1 所有可能的序列多样性所必需的。探针加倍（*N* 个探针，核苷酸的长度为 N）对检测序列多样性是有益的，可以把杂交信号波动的随机性降到最低。如果仅仅依赖于两个突变特异性的探针（每条链一个），产生的波动使信号增强的分析有所偏差。使用信号削弱途径，可以通过绘制测试样品和参比样品与完全匹配的探针杂交信号强度的比率的图，来体现序列的多样性。真正的杂交信号的损失可以通过信号峰清晰的宽度和高度的特征来体现。信号削弱分析的不足之处在于：突变体不易找出；序列变化的特性需由后续地对信号特征削弱的周围片段的双脱氧测序来得到。

3）在片局部测序法：用 4 种荧光标记双脱氧核苷酸（ddNTP），将不标记样品杂交后的芯片上寡核苷酸探针的 3′端用酶延伸作标记。使末端能掺入荧光 ddNTP 的点产生不同荧光，由此来判别序列。

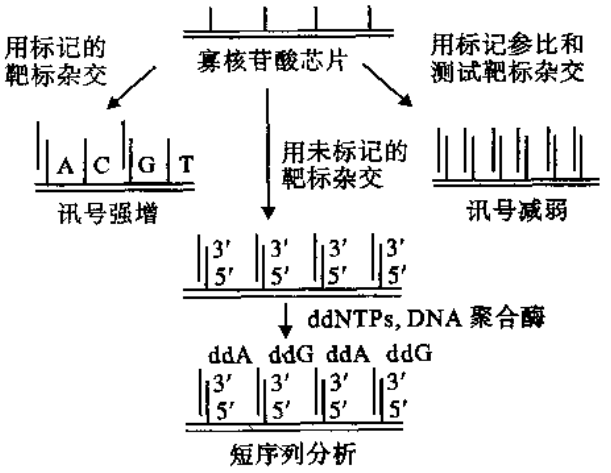


图 17－26　寡核苷酸芯片测序及测变异的 3 种方法

随着人类基因组计划不断深入的研究，以及重要病原体、微生物等基因组测序工作的相继完成，为我们提供了更多可供检测的基因，将促进基因诊断向临床应用的快速发展。可以预料，基因诊断将对未来疾病诊断、未来医疗服务模式乃至医疗保险、社会伦理等产生很大影响。

二、生物芯片在疾病耐药性检测中的应用

基因芯片技术在疾病耐药性检测中的应用体现在以下两个方面：①在肿瘤中，通过检测肿瘤耐药基因的表达变化来分析对药物的抗性；②在感染性疾病中，病原体的耐药性检测可通过两种方式：表达谱芯片检测药物诱导的基因表达改变来分析其耐药性；寡核苷酸芯片检测基因组序列的亚型或突变位点从而分析其耐药性。

目前，包括细菌、酵母菌等许多微生物的基因组序列已基本完成。基因组序列完成为基因组功能分析提供了一个切入点。核苷酸芯片为平行研究大量基因组信息与功能提供了有利的工具。

（一）多药耐药基因的表达检测

肿瘤治疗中对细胞毒性药物的抗性是引起治疗失败的重要原因，是限制化疗的重要因素，机制是复杂的，由肿瘤的综合特征决定，如存活细胞的比例、血液的供给是否充分、特殊的细胞机制及多药耐药表型。多药耐药是指当肿瘤细胞暴露在某一化学治疗药物后会产生对此药及其他结构上没有联系的药物的交叉抗性，可由不同的机制引起，如 *MDR*1、*MRP*、*LRP* 等基因的过度表达，拓扑异构酶 II 和谷胱甘肽代谢的改变等。另外，其他促进 DNA 修复和抑制细胞凋亡的基因表达改变也可能导致多药耐药。

检测多药耐药基因表达的变化不但可以研究恶性肿瘤的不同耐药机制，还可以用于临床诊断，以指导制定治疗方案。目前已建立了几种多药耐药检测方法，在 RNA 水平上有：Northern blot、Slot blot、RT－PCR、RNase protection assay 和原位杂交，从蛋白水平上的检测方法有免疫组化、Western blot 及流式细胞仪等。这些方法一次只能对一个基因进行研究，效率低，难以定量检测耐药基因表达增加的幅度。基因表达谱芯片可同时对成千上万的基因表达进行检测，可以大大加速这方面的研究，在设计芯片时，可以将已知肿瘤相关基因及标记基因都点到芯片上，同时，芯片上还包含目前所有报导过的耐药基因。这样可以同时得到肿瘤的各个方面的信息。另外基因芯片还可以帮助发现新的耐药基因。

（二）病原体耐药性检测

细菌对 3 种以上不同类抗菌药物耐药者即可称为多重耐药菌（multi－drug resistant bacteria，MDR）。MDR 感染在全球的状况十分严重，对婴幼儿、免疫缺陷者和老年人的威胁巨大，1992 年美国疾病控制中心（CDC）的资料表明，有 13 300 例住院患者，是因为对所使用的抗菌药物耐药，细菌感染得不到控制面死亡。MDR 感染已成为治疗上的难点和研究上的热点。

MDR 大多为条件致病菌，革兰阴性杆菌（GNR）占较大比例，如肠杆菌科中的肺炎杆菌、大肠杆菌、阴沟杆菌、粘质沙雷菌、枸橼酸菌属、志贺菌属、沙门菌属等，以及绿脓杆菌、不动杆菌属、流感杆菌等。革兰阳性菌中有甲氧西林耐药葡萄球菌（MRS），尤以 MRSA 和 MRSE 为多；万古霉素耐药肠球菌（VRE），近年来在重症监护室（ICU）中的发病率有明显增高；青霉素耐药肺炎链球菌（PRSP），常引起肺炎、脑膜炎、菌血症和中耳炎，人结核分支菌等。此外尚有淋球菌、脑膜炎球菌、霍乱弧菌等。

耐药性又称抗药性，一般是指病原体的药物反应性降低的一种状态。这是由于长期应用抗菌药，病原体通过产生使药物失活的酶、改变原有代谢过程，而产生的一种使药物效果降低的反应，因而作用的剂量要不断增加。细菌对抗菌药物的耐药机制可有多种，最重要者为灭活酶的产生，如 β－内酰胺酶、氨基糖苷纯化酶等；其次为靶位改变如青霉素结合蛋白（PBPs）的改变等；其他尚有胞膜通透性改变，影响药物的进入；细菌泵出系统增多、增强，以排出已进入细菌内的药物；以及胞膜主动转运减少、建立新代谢途径、增加拮抗药物等，两种以上的机制常可同时启动。耐药菌及 MDR 的发生和发展是抗菌药物广泛应用，特别是无指征滥用的后果。

要解决耐药性这个难题，首先要找到耐药菌的耐药基因，从而根据这些耐药基因设计新型抗生素，或将耐药菌分成不同的亚型，针对不同的亚型在临床上使用相应的抗生素，达到改善治疗效果的目的。国外采用基因芯片技术，检测耐药菌基因的改变，即检测耐药基因。在感染性疾病中，病原体的耐药性检测可通过两种方式：①表达谱芯片检测药物诱导的基因表达改变来分析其耐药性；②寡核苷酸芯片检测基因组序列的亚型或突变位点从而分析其耐药性。

用基因芯片不仅可以同时检测耐药菌的多个耐药基因，还可以同时对多个耐药菌的多个

耐药基因进行检测。对临床上用药和新药物的合成均具有指导作用。

(三)微生物抗药性检测芯片

1.结核分支杆菌耐药性检测芯片。由于利福平、异烟肼、链霉素、乙胺丁醇等12种国际通用抗结核药物长期使用,以及结核病疗程长的因素,自然界和患者体内已产生了对化学药物具有抗性的结核菌株,特别是多重耐药结核病(MDRTB)的流行,增大了治疗难度。临床上选择有效的抗结核药物及提高病人对化学疗法的响应程度要求对药物的敏感性进行测试。目前所采取的敏感性测试需对结核分支杆菌进行培养,而结核分支杆菌培养需要复杂的营养条件,通常耗时5~14 d。

虽然近年来发展了DNA探针,大大减少了试验时间,但仍然存在分支杆菌混合感染的多重诊断问题,特别是那些与HIV-1、2混合感染的病人,迫切需要一种诊断方法鉴定感染分支杆菌种类,鉴定致病菌与非致病菌,鉴定产生抗药性的基因突变,这对流行病学调查与指导临床治疗和愈后具有十分重要的意义。

结核分支杆菌耐药性是由相关基因的突变引起的。Telenti等(1993年)研究首次证明结核分支杆菌对利福平的抗药性与RNA聚合酶β亚基(*rpoB*基因)有关,约90%~95%的耐药菌株在*rpoB*基因3 534 bp编码区中有一81 bp突变节段;在该基因的511~533位密码子上碱基的替换、缺失、插入已有相关报道。

亚能生物技术(深圳)有限公司开发的第一代产品TBChip™-r专为快速、简便地检测利福平耐药性突变株而设计的。这是一种基于DNA芯片技术识别结核杆菌耐药性基因型的新技术。这个过程包括四个主要步骤:从样品DNA中扩增含有突变位点的特异DNA片断(靶DNA),用荧光染料标记靶DNA,荧光标记的靶DNA与TBChip™-r上的已知的DNA阵列杂交,最终获取图像并进行数据分析。整个检测周期为3 d。这项技术为临床基因检测提供了更为简单、快捷的方法。利福平耐药性是多重耐药结核病的潜伏性标志,TBChip™可以快速的检测出多重耐药结核菌(MDRTB);TBChip™-r也可以检测出已知的绝大多数(>90%)的突变型。

异烟肼(isoniazid, INH)是治疗肺结核的常用药物,其治疗结核病的机制是它阻断了分支菌酸的生物合成途径。Wilson等人根据已测序的肺结核分支杆菌基因组序列,用PCR方法扩增了3 834个ORF(占全部ORF的97%),固化在玻片上,制成检测肺结核杆菌基因表达的DNA芯片。用INH处理敏感菌株,发现除了生化途径已清楚的与分支菌酸合成相关的一些基因转录水平发生变化外,还发现*EfpA*基因的表达也被诱导发生了变化,推测*EfpA*基因也参与了分支菌酸的生物合成,而*EfpA*基因只在分支菌属中一些致病的种类中才存在,故*EfpA*基因可以作为治疗结核病的新靶点。另外,分别用INH和Ethionamide处理敏感菌株,获得了相似的基因表达谱,证实了两者具有相同的作用机制。又如,Wilson还使用包含有肺结核杆菌基因组*ORF*的97%的序列的基因芯片,对应用抗结核杆菌药物异烟肼诱导前后表达的变化,结果证明肺结核杆菌中脂肪酸合成酶Ⅱ、*FbpC*、*EfpA*、*fadE23*、*fadE24*和*ahpC*基因发生改变与耐药性有关,并为新药物作用的靶目标研究及指导抑制这些靶目标试剂和药物的合成提供指导。如图17-27。

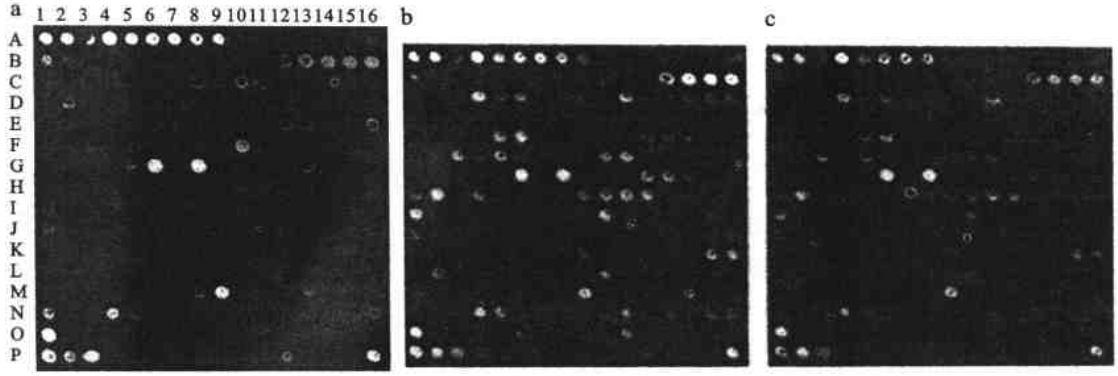


图 17－27　微阵列杂交分析监测异烟肼诱导的 mRNA 表达情况

(a) 含有 203 个不同结核分支杆菌开放阅读框对应斑点的子阵列显示由异烟肼诱导的基因表达情况，包含了被认为是编码分支菌酸成分的基因。伪彩色合成图像为异烟肼处理 4 h 的表达情况。根据荧光强度两条通道为伪彩色：红色(异烟肼治疗)、绿色(非异烟肼治疗)。在 B13－16 位置占优势的红斑对应的是 *FAS*－Ⅱ基因簇的基因(*Rv*2244－7)；P2－3 相对应的斑点是预先用 Cy3(P2)或 Cy5(P3)标记的 DNA 阳性对照。对照斑点含有结核分支杆菌全基因组的 DNA(A1、P1 和 P16)和核糖体 DNA(A4－10)。(b) 用异烟肼处理异烟肼耐药、乙硫异烟胺敏感菌株(4309A)揭示在相应的 mRNA 水平上未显示明显的变化。(c)用乙硫异烟胺处理相同菌株产生的反应形式与用异烟肼处理异烟肼敏感株所产生的反应形式相同(a)。

利用基因芯片探针杂交方法可以对 2 个以上不同抗药性基因同时进行分析，如对结核分支杆菌异烟肼、对氨基苯甲酸、异胺丁醇等多种药物的抗药性基因 *kasa*、*pncA*、*emb* 等都可以同时进行监测，包括多个结核分支杆菌抗药性基因的探针芯片现在已经得到应用。Head 等(1999)在 PCR 后 1 h 内应用中等密度芯片对 *rpoB* 基因序列扫描，得到了肉眼可见的诊断结果，对抗药性菌株鉴定准确、快速。

2.肝炎病毒抗药性检测芯片。肝炎是一种世界性传染病，中国是高发区之一，最常见有甲、乙、丙、丁、戊及庚型肝炎。其中乙、丙、庚型肝炎病毒的感染率最高，且与慢性肝炎、肝硬化、肝细胞癌变密切相关。为遴选抗病毒药物，评价临床疗效，除生化、免疫和病理指标外，还应明确病毒的核酸结构和体液病毒核酸的含量。用 DNA 测序、PCR 等方法虽可达到目的，但由于技术难度大、实验成本高而限制了应用。目前在乙型肝炎的抗病毒治疗中，基因工程干扰素是公认的治疗慢性乙型肝炎病毒(HBV)感染的有效药物之一，但因其有效适应症少，远期有效率低和有一定的不良反应，而无法成为治疗 HBV 感染的主要药物。HBV 复制时有一个逆转录过程，对逆转录酶有抑制作用的药物就有可能抑制 HBV 的复制，进而发挥抑制 HBV 的作用。近年来在艾滋病的研究中发现了许多新的核苷类抗病毒药物对 HBV 有抑制作用，其中应用拉咪呋啶(Lamifudine)治疗慢性乙型肝炎的报告最多。目前国外已基本完成了Ⅰ、Ⅱ、Ⅲ期临床试验，累计病例已达 30 000 例以上。美国已于今年批准上市。我国自 1996 年起已在北京、上海和广州联合进行拉咪呋啶治疗慢性乙型肝炎的临床研究。

拉咪呋啶是一种胞嘧啶核苷衍生物(3′硫胞嘧啶核苷左旋异构体)，属于核苷类化合物，为口服的 2′3′双脱氧核苷酸。通过其代谢产生 5′三磷酸活性产物而发挥作用。对 HIV 和 HBC 的逆转录酶的活性具有显著的抑制作用，主要抑制 DNA 聚合酶形成负链引物的过程，从而抑制 HBV 的复制。无论是短期还是长期的应用，均获得了显著的抑制病毒复制，成为目前抗乙型肝炎病毒的主要临床用药。但随着使用时间的延长，发现在不少病人中病毒复制再度活跃，出现明显的耐药。而出现耐药的主要机制是乙肝病毒 DNA 聚合酶活性区 YMDD(酪氨酸、蛋

氨酸、门冬氨酸)发生变异,产生 HBV 耐药突变株。经对这些出现耐药的毒株进行测序研究,1996 年首次发现并证实耐药与病毒的 YMDD 区的基因突变有关。其后,不少学者又对发现的位点,进行了耐药的确证研究。应用特定序列的寡核苷酸为探针,采用基因芯片技术,可检测出最常见的 4 个突变位点。Allen Niesterst 等将 YMDD 区的 M552V、M552I、L528M 位点的突变株构建成质粒,转入 2215 细胞内,拉咪呋啶抑制病毒复制的作用显著减弱万倍以上,在体外证实了其耐药与上述位点突变的关系。随后,又有越来越多的耐药毒株被发现。这些突变株通过诱导或治疗选择下成为优势毒株从而导致耐药。近来临床流行病学研究发现,应用一年后的耐药发生率高达 17% ~ 30%。国内一组最近的报告亦显示,其耐药的发生率在 52 周时为 43.37%(36/83),104 周(2 年)时的突变发生率为 48%(36/75)。

尽管从出现耐药毒株的基因序列分析中发现,与拉咪呋啶耐药相关的突变位点高达 28 个,但在现有文献报道中,上述位点出现的突变,均是伴随着 YMDD 区的突变而共同存在,而其他位点在耐药中的重要性及特异性,尚待进一步证实。

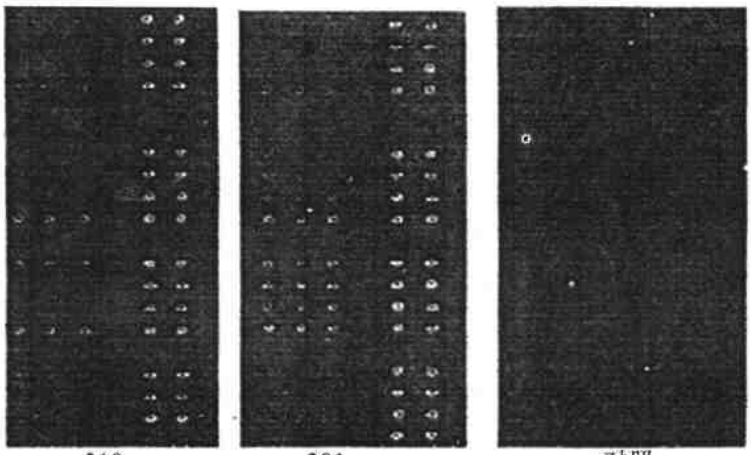
拉咪呋啶(3TC)抗 HBV 治疗中出现的 YMDD 变异往往发生于应用 3TC 半年后。作用靶位为病毒的 DNA 聚合酶(DNAP),抑制其活性,同时易引起 DNAP 的多位点突变。基因芯片的高通量、平行检测技术针对引起 YMDD 变异的众多基因突变位点设计探针,将之结合于同一张芯片或与上述分型基因芯片合并,从血清样本中抽取病毒 DNA,经体外扩增后与芯片探针杂交,依据杂交信号判定 HBV 的 YMDD 变异,得出是否产生耐药性的结论。

YMDD 变异过去常用 PCR 直接测序或快速 PCR 等方法进行检测。最近,上海联合基因集团博华基因芯片技术有限公司开发了一种检测 YMDD 变异的基因芯片,它是选用 HBV 的 *P* 基因的保守区部分序列,设计相关引物,扩增出包含有 YMDD 变异区的核苷酸片段,针对 L528M、A548V、M552V 等几个位点设计探针,将探针固定在特殊处理的玻片上。利用它可一次同时检测多个突变位点,同时测定变异株和野生株。对于临床检测拉咪呋啶治疗后 HBV YMDD 变异是有价值的。初步临床检测,其检测灵敏度约 10^4 copies/mL,检测结果与测序结果比较,符合率为 100%,并且能够检测到野生与突变共生的类型(测序做不到)。图 17 - 28 显示的是 20 份血清样品经拉咪呋啶乙肝耐药基因芯片的检测,其中 6 份为野生型,其余 14 份为变异毒株。

实验研究和临床试验都证明拉咪呋啶可迅速降低 HBV - DNA 的浓度,改善肝脏组织的病变。是一种治疗乙型肝炎病毒感染很有希望的新药。拉咪呋啶治疗乙型肝炎的适当疗程至今还在探索。由于拉咪呋啶的作用机制是抑制而不是杀灭 HBV,肝细胞内的 HBV 能否消除主要取决于机体免疫清除能力,因而当停药后病毒复制仍可恢复。临床实践也证明,许多患者停药后出现"反跳"现象,肝炎复发。

乙型肝炎病毒突变检测基因芯片(microarray genechip)是专门针对服用拉咪呋啶治疗慢性乙肝的患者而生产的,它利用基因芯片这一新型的高通量、平行的基因分析技术,合理设计基因矩阵,可以一次检测多个乙肝病毒相关位点的突变,判断野生株和突变株的共生状态及相对优势株,具有特异性强、灵敏度高、快速、方便等优点。我国迈科锐乙肝病毒突变检测基因芯片选取最常见的三个突变位点,能检测出 95%以上突变毒株;还可以根据临床需要或新位点的发现,随时增加新的检测位点,为临床医生及时发现病毒变异,及时调整用药方案提供有效地帮助。亚能生物技术(深圳)有限公司开发的 HepVchip™是一种基于 DNA 芯片技术快速检测肝炎病毒的型别和耐药突变基因型的新技术,可对肝炎病毒的分型、变异、突变和病毒核酸含量进行检测。为临床的准确诊断、合理用药、判定预后提供切实可信的依据。HepVchip™能够对所有的肝炎病毒的分型、变异、突变和病毒核酸含量进行高通量、平行检测;还可快速的检测

拉咪呋啶抗药性乙型肝炎病毒。



219
(共生型)　　201
(YIDD)　　对照
(阴性)

血清编号	基因芯片法	DNA 测序法	快速 PCR 法
201	YYDD,528 突变	YYDD,528 突变	YYDD
205	PCR 阴性	野生株	野生株
209	野生株	野生株	野生株
221	YIDD	YIDD	YIDD
229	YYDD,528 突变	YYDD,528 突变	YYDD
249	YYDD	YYDD,528 突变	YYDD
250	YYDD,528 突变	YYDD,528 突变	YYDD
256	YYDD,528 突变	YYDD,528 突变	YYDD
257	YYDD,YIDD	YIDD	YIDD
262	野生型	野生型	野生型
183	YYDD,528 突变	野生型	YYDD,YIDD
185	YYDD 和野生株共生,528 突变	野生型	YYDD,YIDD
189	YYDD 和野生株共生	野生型	YYDD
194	野生型	野生型	野生型
202	YYDD 和野生株共生,528 突变	野生型	YYDD,YIDD
214	YYDD 和野生株共生,YIDD	YIDD	YYDD,YIDD
218	野生型	野生型	野生型
231	野生型	野生型	野生型
243	YYDD 和野生株共生,YIDD	野生型,528 突变	YYDD,YIDD
248	YYDD,528 突变	YYDD,528 突变	YYDD

图 17－28　拉咪呋啶乙肝耐药基因芯片检测

用基因芯片不仅可以同时检测耐药菌的多个耐药基因,还可以同时对多个耐药菌的多个耐药基因进行检测。对临床上用药和新药物的合成均具有指导作用。

3. 性传播疾病病原体检测芯片。性传播疾病(sexually transmitted disease, STD)在世界范围内广泛流行,是已扩展至今至少包括 50 种微生物所感染的疾病。其中难以根除的主要原因是

其病原体发生基因变异而具有抗药性。艾滋病(AIDS)是由艾滋病病毒(HIV)感染所致。感染后可以完全无表面症状,潜伏期可达七年,病发可导致死亡。现用于治疗 AIDS 的药物主要是病毒逆转录酶 RT 和蛋白酶 PRO 的抑制剂,但在用药 3～12 月后常出现耐药,其原因是 *rt*、*pro* 基因产生一个或多个点突变。*rt* 基因四个常见突变位点是 Asp67→Asn、Lys70→Arg、Thr215→Phe、Tyr 和 Lys219→Glu,四个位点均突变较单一位点突变后对药物的耐受能力成百倍增加。如将这些基因突变部位的全部序列构建为 DNA 芯片,则可快速地检测病人是这一个或那一个或多个基因发生突变,从而可对症下药,所以对指导治疗和预后有很大的意义。

淋病由淋病双球菌 (Neisseria gonorrhoea)感染所引致。五成女性及两成男性不会出现任何症状,或症状轻微面被忽略,严重者可引导不育。目前已经发现有红霉素、奎诺隆类抗生素、环酶素等抗药性突变株。梅毒(Syphilis)由梅毒螺旋体感染所引致。在常见性病中占第五位,潜伏期为 9～90 d。发病后病菌入侵身体心脏、眼睛、骨骼及脑部等各部分器官,将造成失明、心脏病及精神病等多种疾病,严重者可导致死亡。其病菌基因的变异产生了红霉素抗药性。

性传播疾病检测型基因芯片是利用核酸分子碱基互补配对原理,将多种病原体特异基因序列提取以后,按微点阵方式固定于芯片上,使其与标记的样品核酸分子杂交,通过检测杂交信号强度,获得样品核酸分子的数量和序列信息,从而能在一份样品上同时检测多种性传播疾病的病原微生物。该芯片具有灵敏度、特异性高,操作简便、快速,自动化程度高,结果客观性强等特点,弥补了传统技术的不足。据介绍,这种基因芯片在国内尚属首例。

我国自行设计的 STD 检测基因芯片,与传统的检测方法相比,STD 基因芯片具有灵敏度、特异性高,假阳性率、假阴性率低,操作简便、快速,自动化程度高,结果客观性强的特点。可一次性进行多种性病的检测,可广泛应用于涉及性传播疾病诊断的各种领域。

亚能生物技术(深圳)有限公司开发的另一产品 STDChip™是一种基于 DNA 芯片技术识别性传播疾病病原体基因型的新技术,专为快速、简便地检测不同病原体及其耐药性突变株而设计的。可以辅助医疗机构对性病的控制;为有病征或已知有病者作进一步确认及评估;进行性病治疗的效果评估和婚前男女双方检查。

如果要一次性同步诊断,多基因同步检测,同步定性及定量分析,只要有完整配套设备及软件,可用于:①泌尿生殖道疾病病原微生物的诊断和鉴别诊断;②泌尿生殖道疾病治疗效果的评价及疗效观察;③泌尿生殖道感染微生物的基因结构,基因表达、基因变异、基因定量的研究;④药物作用靶位点研究。

三、前景与展望

台湾有关科研机构研发的"抗药性基因芯片",已经投入生产。据介绍,这种基因芯片当天就能让抗药性病原体现形,应用这种技术,遭病原菌感染的患者就不必再苦候数周等待细菌培养结果了。

目前"抗药性基因芯片"主要针对几种常见抗药性病原菌而设计的,除了上述的几种病原体外,还包括肺炎双球菌、金黄色葡萄球菌等。专家说,在抗药性问题日趋严重的情况下,这种基因芯片可望协助临床医生快速掌握患者情况,从而给与予有效的药物治疗。

除用于临床检验外,"抗药性基因芯片"也是药厂研发治疗抗药性细菌感染新药的重要工具。它可帮助研究人员了解抗药性细菌对新药的反应,从而加快研发速度。

"抗药性基因芯片"的研发除了需要芯片制作的技术平台之外,找出具有标记性的基因片段也是重要工作。专家说,这种基因片段就像是抗药性细菌的身份证,只要检出就表明抗药性

病菌已经存在。

目前，国内外已经有许多从事 DNA 芯片研究的公司，每家公司的芯片技术都各具特色，应用目的也不尽相同。但要成为实验室研究或临床可以普遍采用的技术仍有一些关键问题亟待解决如：①基因芯片的特异性的提高；②样品制备和标记操作的简化；③增加信号检测的灵敏度；④高度集成化样品制备、基因扩增、核酸标记及检测仪器的研制和开发。

近年兴起的 DNA 芯片技术有可能给医学微生物学带来一次新的飞跃。DNA 芯片技术的最大优点为可快速、准确地提供病原体的遗传信息。在医学微生物学中已应用于以下几个方面：微生物基因组及后基因组的研究；可分析基因序列，研究基因的转录表达；微生物感染的快速诊断，可同时检测多种病原体基因；微生物变异，耐药机制的研究；微生物基因分型及分子流行病学的调查；抗微生物感染药物的研制等等。目前存在的主要问题是如何改进 DNA 芯片的制造工艺，并使标本制备、检测设备等简化，使 DNA 芯片技术更简便、可靠、价廉。

（廖　芳）

第七节　基因芯片技术在药学方面的应用

生物芯片已成为高效、大规模获取相关生物信息的重要手段，是被公认为将给 21 世纪生命科学研究带来一场革命的高新技术。它具有高度平行性、多样性、微型化、自动化、成本低的特点，在药学领域有着广泛的应用。

一、生物芯片在药物靶点发现与药物作用机制研究中的应用

在当代药物开发过程中发现和选择合适的药物靶点是药物开发的第一步，也是药物筛选及药物定向合成的关键因素之一。在药物靶点发现和药物作用机制研究中所使用的生物芯片主要是指基因芯片。应用基因芯片可以在基因水平上寻找药物作用的靶标。基因芯片可以平行测定几千个基因的表达方式，以发现有意义的靶标。

基因芯片在药物靶点发现与药物作用机制研究中的应用具体表现在以下几个方面：

（一）比较正常不同组织细胞中基因的表达模式

基因的表达模式给它的功能提供了间接的信息。例如只在肾脏中表达的基因就不大可能与精神分裂症有关。一些药物的靶点是在整个身体中分布广泛的蛋白质，这类药物的不良反应往往比较大。而选择只在特异组织中才表达的蛋白作为药物筛选的靶点，可以减少药物对整体产生的不良反应，因而更引起人们的关注。例如骨质疏松症（osteoporosis）与破骨细胞（osteoclasis）的功能有关，破骨细胞可以破坏并吸收骨质，当骨质的形成与破坏出现不平衡的时候，就会导致骨质疏松症。如果破骨细胞的功能得到抑制，那么就可以控制骨质疏松症的发生和发展。利用已有的人类 ESTs 序列和 DNA 芯片技术，可以容易地得到只在破骨细胞中进行表达的基因如 *cathepsink* 基因，它编码半胱氨酸蛋白酶。以 *cathepsink* 基因作为靶标，筛选对它有抑制作用的药物，就有可能得到治疗骨质疏松症的药物。但是这种方法也有其局限性，它只能得到 mRNA 水平的表达谱，另外组织一般由多种细胞组成，面要将这些细胞分离很困难。

（二）研究正常组织与病理组织基因表达差异

正常组织在病变的过程中，往往伴随着基因表达模式的变化。基因表达水平的升高或降低，可能是病变的原因，也可能是病变的结果。若基因表达的变化是病变的原因，则以此基因为靶点的药物就可能逆转病变；若基因表达的变化是病变的结果，则以此基因为靶点的药物就

可能减轻病变的症状。DNA 芯片技术可以在病理组织与正常组织之间一次比较成千上万个基因的表达变化，找出病理组织中表达异常的基因。Heller 等提取正常及诱发病变的巨噬细胞、软骨细胞系、原代软骨细胞和滑膜细胞的 mRNA，用包含细胞因子、趋化因子、DNA 结合蛋白及基质降解金属蛋白酶等几大类基因的 cDNA 芯片进行筛选，发现了数种变化明显的基因。其中除了有已知与类风湿关节炎有关的 *TNF*，*IL*－1，*IL*－6，*IL*－8，*G*－*CSF*，*RANTES*，*VCAM* 的基因外，还有编码基质金属弹性蛋白酶 *HME*，*IL*－3，*ICE*，趋化因子 *Groα* 等的基因。而以前认为金属弹性蛋白酶只存在于肺泡巨噬细胞和胎盘细胞中。弹性蛋白酶可以破坏胶原纤维及组织基底膜层，它在类风湿关节炎病理组织中的出现，为治疗该病提供了新的药物靶点。

Kapp 等用包含 950 个 DNA 探针的 DNA 芯片分析比较霍奇金病细胞系 L428 及 KMH2 与 EB 病毒永生化的 B 淋巴细胞系 LGL－GK 的基因表达谱发现霍奇金病源的细胞系中白细胞介素－13(*IL*－13)及白细胞介素－5(*IL*－5)表达异常增高；用 IL－13 抗体处理霍奇金病源的细胞系可显著抑制其增殖。此发现提示，IL－13 可能以自分泌形式促进霍奇金病相关细胞增殖。IL－13 及其信号转导途径可能成为霍奇金病治疗及药物筛选的新靶点。

(三)建立模式生物细胞中的基因表达模型

采用模式生物细胞进行试验，条件容易控制，对模式生物基因表达的研究将启发人们发现和确认新的药物作用靶点。目前，已有多种模式生物(如酵母)的基因组计划已经完成。酿酒酵母(saccharomvces cerevisiae)就是一种可用来进行药物筛选的较为理想的模式生物。它是真核生物而且基因组已全部测序，细胞繁殖快，易于培养，与哺乳动物细胞有许多共同的生化机制。现在已经发现，在酵母细胞中存在许多与人类疾病相关的基因。如人类 Werner's 综合征表现出早熟的特征，其细胞的生活周期变短。人类与此疾病相关的基因与酵母中编码 DNA 解旋酶的 *SGS*1 基因极为相似。Botstein 等得到了 *SGS*1 基因突变的酵母菌株，此突变菌株生活周期变短，细胞的表型特征与患有 Werner's 综合征的细胞相似。已有一些研究小组根据公布的酵母基因组序列，用 PCR 方法扩增了酵母 6 000 个开放阅读框(open reading frame，ORF)片段，制成 DNA 芯片，在整个基因组的范围内对酵母的基因表达进行检测。

(四)建立病原体基因的表达模型

(见本章第六节微生物抗药性检测芯片)

(五)研究药物处理细胞后基因表达变化

药物与细胞(特别是敏感细胞)相互作用，将引起细胞外部形态及内部正常代谢过程的一系列变化。其内部生理活动的变化可集中表现在其基因表达的变化上。通过测定分析药物对细胞的基因表达的影响，可推测药物的作用机制，评价药物活性及毒性，进而确证药物靶点或者发现新的药物靶点。通过 DNA 芯片测定药物诱导的细胞基因表达变化来进行药物筛选与研究，对那些用常规方法很难追踪监测的药物或需要很长时间才能得到药物临床实验结果时，显得尤为有用。

无论是从药物的治疗效果还是从经济效益方面来看，药物进入市场前必须对其进行严格的安全性检测，这样做的前提是对药物的作用机理必须十分了解。根据分子药理学的研究进展，已知药物的作用会使细胞内某些特定基因的表达丰度发生相当不同程度的变化，因此可以利用 cDNA 表达谱的改变来寻找药物作用的靶基因。

通过监测阳性药物处理前后组织细胞基因表达变化情况可以获得许多十分有价值的信息。首先，经药物处理后表达明显改变的基因往往与发病过程及药物作用途径密切相关，很可

能是药物作用的靶点或继发事件，可作为进一步药物筛选的靶点或对已有的靶点进行验证；其次，药物处理后基因表达的改变对药物作用机制研究有一定的提示作用。

理论上讲，药物作用诱导的细胞基因表达变化应与缺失编码该药物作用靶点的基因引起的基因表达变化相似。现以目前临床上最常用的免疫抑制药物环孢素 A(Cyclosprin A，CsA)和他克洛司(Tacrolimus，FK506)为例说明基因分析在药理研究中的重要性。

为了筛选 CsA 和 FK506 发挥免疫调节作用的特异性靶基因，或者验证某些一度被认为是 CsA 和 FK506 作用标志基因的表达是否真与该药物的作用相一致，就必须建立一个高保真的、高通量的基因表达分析技术，还需要生物信息数据库的高技术平台的支持。基因芯片技术与 Gene Logic 数据库的结合无疑是一条非常有力的分析途径。

Marton 等使用含有 6 065 个酿酒酵母的 ORF 的 cDNA 芯片检测发现，由免疫抑制剂 FK506 诱导的基因表达变化在缺失编码被 FK506 抑制的蛋白的基因变种中未观察到，但在缺失与 FK506 作用无关的基因变种中却观察到了。于是推测 FK506 除了与免疫亲合蛋白(immunophilins)结合外还有其他被忽略的作用靶(off－target)。在实验基础上，他们提出了所谓的“解码器策略”(decoder strategy)，即首先比较经过药物作用的野生株和缺失某些基因的变种株的基因表达谱，若二者相似，则将这种变种挑选出来，并将其与药物作用。如果突变基因所编码的蛋白质参与的生物途径受药物的影响，则药物诱导该变种的基因表达变化与药物诱导野生株基因表达变化将不同(见图 17－29)。用这种策略不仅可以确证药物作用靶，还可以发现未曾引起人们注意的作用靶，并可从这些被忽略掉的靶中推测药物的毒副作用。

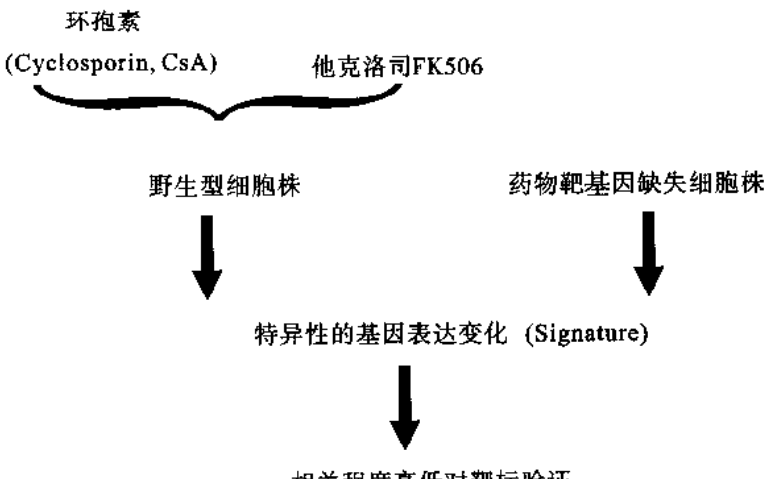


图 17－29　FK506 和 CsA 靶标检测的“解码器”策略

CsA 和 FK506 是结构不同的两种免疫抑制药物，它们分别与亲环素(Cyclophilin，CyP)及 FK506 结合蛋白(FKBP)结合。CyP 与 FKBP 具有脯氨酸异构酶的活性，当它们与 CsA 和 FK506 结合后酶的活性将受到抑制。CsA·CyP 及 FK506·FKBP 的复合物还可抑制依赖于钙的丝氨酸/苏氨酸磷酸酶的活性，阻碍 NFAT(nuclear factor of activated T cell)活化，从而达到抑制某些基因的转录活化(见图 17－30)。

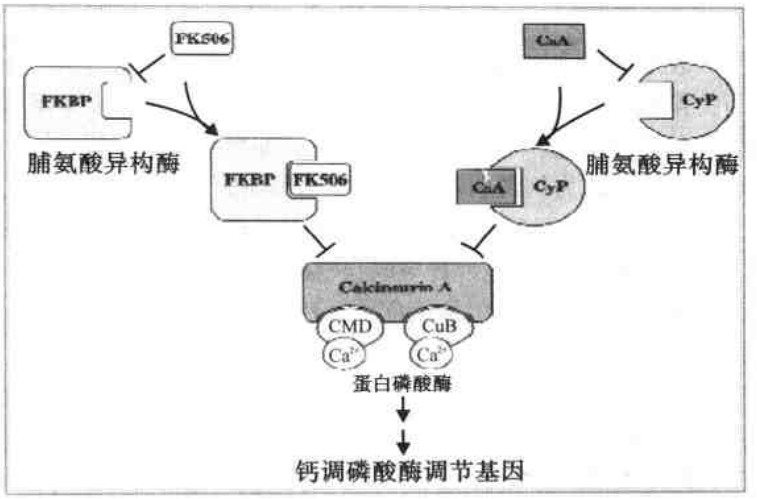


图 17－30　FK506 和 CsA 作用的分子靶标

首先将正常人外周血淋巴细胞收集，分离其 T 淋巴细胞，用抗 CD3 抗体和 PMA 进行刺激，激活后的 T 淋巴细胞在加入 CsA 和 FK506 前、后分别提取 mRNA，用 Cy3 和 Cy5 荧光素标记制备 cDNA 探针。然后将这些探针与固定了上千万个人的 cDNA 表达谱芯片进行杂交。通过杂交信号分析，比较 CsA 和 FK506 作用前后基因表达水平的差异，从中初步筛选出那些与 CsA 和 FK506 作用相关的基因，再根据相关程度的高低可进一步确定主要的靶基因和靶基因相关基因。在此初步筛选的基础上，接下来的工作就是逆向确定筛选出的靶基因是否真的与 CsA 和 FK506 作用特异相关，或者说能否通过这些基因的表达水平来预测药物作用的效果。从 CsA 和 FK506 处理过的正常的活化 T 细胞 mRNA 中制备 cDNA 探针，同时还从来源于待验证靶基因缺失(或突变)的 T 细胞株的 mRNA 中制备 cDNA 探针，将它们分别与 cDNA 表达谱芯片杂交，以比较基因表达差异。图 17－31 的实验结果显示用 FK506 处理的正常激活的 T 细胞与已知靶基因 *Calcineurin* 发生突变的 T 细胞株的基因表达谱存在相似性($\gamma = 0.75 \pm 0.03$)，证明 FK506 对 T 细胞作用的确是通过靶基因 *Calcineurin* 来实现的(图 17－31)。

近年来的研究还表明，CsA 和 FK506 发挥免疫抑制作用除了依赖于钙调节磷酸酶－NFAT 途径外，还可通过非钙调节磷酸酶途径，如 JNK(C－Jun NH2－terminal kinase)及 P38 途径，因而探索 CsA 和 FK506 作用的分子机理还需要不断的深入和发掘，要做到这点，基因芯片分析技术将大有作为。

Cyclin 依赖型激酶(cyclin－dependent kinases，CDKs)在细胞增殖中有着重要的作用，是一种抗肿瘤药物的靶点。Gray 等从 2,6,9,－三取代嘌呤化合物库中筛选 CDKs 的抑制剂。他们将体外有活性的 2 种 flavopiridol 化合物和 52 号(代号)化合物与酵母作用，然后用共含有260 000 个长度为 25 个碱基的一组寡聚核苷酸芯片检测酵母基因表达变化。试验结果表明，几乎所有的已知与细胞周期相关的基因表达下调。虽然 flavopiridol 和 52 号化合物体外活性相似，但他们引起的酵母基因表达的变化却有显著的差异，表明二者对酵母的作用途径是不一样的。

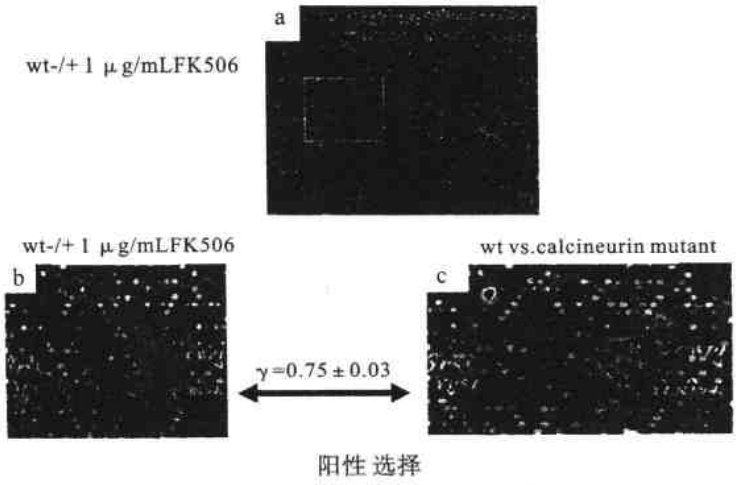


图 17－31　FK506 处理的野生型与 cna 突变株基因表达相似程度的比较

应用 DNA 芯片还可直接筛选特定的基因文库以寻找药物的作用靶点。如给酵母某一特定的呈单倍体状态的基因对应的位置上放置一个遗传标记，该标记可被 DNA 芯片所识别，那么通过比较药物作用前后用芯片检测整个文库的结果，便有可能获得药物作用的靶基因。研究者称此种筛选方式为 haploinsufficiency drug screen。

除了基因芯片外，组织芯片、蛋白质芯片和细胞芯片也在药物研究中崭露头角。最近，耶鲁大学的研究小组首次报道了真核生物蛋白质组水平的蛋白质微阵列芯片。他们表达和纯化了酵母的 5 800 种蛋白质，并将这些蛋白质点样固定在载玻片上，制作了酵母蛋白质组微阵列芯片。他们使用这种芯片筛选能与特定蛋白质和磷脂相互作用的蛋白质，发现了新的能与钙调蛋白和磷脂相互作用的蛋白质。这种蛋白质微阵列芯片可以用于筛选与蛋白质相互作用的药物，还可以用于检测蛋白质翻译后的修饰。他们的研究成果证实了制作和使用蛋白质组微阵列芯片进行功能分析检测的可行性，并向人们预示了蛋白质微阵列芯片在药物开发领域的广阔应用前景。

Zlauddin 和 Sabatinl 发明了一种细胞微阵列芯片。他们首先将不同的质粒 DNA 点在玻璃片上，做成质粒 DNA 微阵列芯片。接着用脂质转染试剂处理该质粒 DNA 微阵列芯片，然后在处理好的质粒 DNA 微阵列上培养哺乳动物细胞。点在芯片上的质粒 DNA 在转染试剂的帮助下原位转染哺乳动物细胞，在质粒 DNA 微阵列的每一个质粒样品点的相同位置形成了转染了该质粒的细胞群。细胞因获得了外源 DNA 而获得了新的性状。这样，由质粒 DNA 芯片制成了由不同性状细胞组成的细胞微阵列芯片。他们尝试用这种细胞芯片来确证药物作用靶点，寻找能改变细胞生理状态的基因产物。这种新型细胞芯片可以用来在哺乳动物细胞内高通量筛选，有可能成为候选先导分子的化合物、蛋白质或寡核苷酸。在功能基因组研究和药物开发等领域具有很大的应用潜力。

二、生物芯片在毒理学研究中的应用

传统的毒理学研究是通过大量的动物实验来进行，而慢性毒性和长期毒性一般是通过新药的Ⅲ期临床试验来观察的。由于慢性毒性和长期毒性难以在短时间内观察出来，往往会造成一些严重的后果。基因芯片技术可将药物毒性与基因表达特征联系起来，通过基因表达分

析便可确定药物毒性，使得药物毒性或不期望出现的效应在临床实验前得以确认。用基因芯片可以在一个实验中同时对成千上万个基因的表达情况进行分析，为研究化学或药物分子对生物系统的作用提供全新的线索。该技术可对单个或多个有害物质进行分析，确定化学物质在低剂量条件下的毒性，分析、推断有毒物质对不同生物的毒性可比性。基因芯片技术应用于药物毒副作用，改变了传统的毒理学研究模式，节省大量的动物试验，提高用药的安全性。

已经有不少研究工作表明，利用 DNA 芯片预测化合物毒性和对毒性物质进行分类是可行的。Waring 等用 15 种已知的肝毒性化合物处理大鼠。这些毒物将对肝细胞造成多种伤害，如 DNA 损伤、肝硬化、肝坏死和诱发肝癌等。从大鼠肝脏中提取 RNA，用 DNA 芯片作基因表达分析。通过将基因表达结果与组织病理分析和临床化学分析的结果进行比较，发现两者有很强的相关性。该结果表明，DNA 芯片分析是一种可以用来分析药物安全性和对环境毒物进行分类的灵敏度较高的方法。在另一报道中他们用同样的 15 种化合物作用大鼠的肝细胞，再用 DNA 芯片作基因表达分析，结果显示具有相似毒性机制的化合物所获得基因表达谱具有相似性。Gerhold 等给大鼠服用苯巴比妥和地塞米松等药物，使用寡核苷酸芯片，检测了大鼠肝组织中与药物代谢、毒性和能量代谢相关基因的表达。最后，通过分析基因表达变化的结果就可以推测药物代谢与毒性的情况。Bartosiewicz 等则利用 DNA 芯片对环境毒物进行了检测。

目前已有多种较为成熟的毒理学 DNA 芯片相继问世。美国国立环境卫生研究院分子致癌机制实验室的 Barrett 等人研制了一种名为 ToxChip 的 DNA 芯片，可以灵敏的检测有害化学物质对人体基因表达的作用；Gene Logic 公司的产品 Flow - thru Chip 已经试投入商业运用，可用以检测药物和毒物对生物体的影响，他们还建立了庞大的基因表达数据库，可以用于药物靶点确认和毒性预测；Syngenta 公司和 AstraZeneca Pharmaceuticals 公司的科学家设计制作了被称为 ToxBlot arrays 的 DNA 芯片，含有大约 13 000 个人的基因，包含了所有毒理学家感兴趣的基因家族和信号通路相关基因；美国化学工业毒物研究所中专门有一个工作小组用微阵列技术研究一些致癌毒物对人体的作用。

三、生物芯片作为超高通量药物筛选平台的应用

传统的新化合物筛选是以分子功能分析和改进为主要手段，通过大量化合物合成与结合实验来确定可能有效的新药，这是一种低效率的过程。近年来，根据组合化学理论设计的各种化合物文库开始用于药物筛选，以及高通量筛选（high throughput screening，HTS）技术的出现，意味着可以在很短的时间内对大批量的化合物进行筛选。

组合化学是采用一系列小分子单元如氨基酸、糖和核酸等，通过化学合成或生物合成的步骤获得结构相近却又互不相同的大量小分子组成化合物文库。化合物文库包括化学文库和生物文库。化学文库是指由化学合成的大量小分子物质组成的文库。生物文库是相对化学文库而言的，其最主要的特点是可以进行扩增和放大，并通过 DNA 测序来对筛选到的克隆分子进行鉴定。组合化学技术的出现使得快速合成多种系列化合物成为可能。

应用生物芯片技术对生物文库进行药物筛选，其基本原理是在芯片或合成的微球珠等固相表面上进行原位文库合成和筛选。现已有针对不同类型的药物而制作的芯片，如筛选核酸类药物的 DNA 芯片，筛选肽或蛋白质药物的微集阵列芯片，筛选抗原、抗体的免疫芯片等。

提高药物筛选的通量，实现超高通量筛选有 2 条途径：一是微型化，一是自动化。生物芯片作为一种新型技术平台，正可满足超高通量筛选微型化和自动化的需要。生物芯片技术应用于超高通量筛选有 2 个发展方向：一是微孔板 / 微阵列技术，一是微流体芯片技术。

(一)微孔板/微阵列技术

相当多的高通量筛选方法是以微孔板为基础的。微孔板技术的发展主要表现在板孔数的增加。目前,使用最多的是96孔及384孔板。Oldenburg等报道了用9 600孔板(0.2 μL/孔)分析系统,以金属蛋白酶为靶,对复合或分离纯化的化合物库进行筛选。虽然随着材料科学和加工技术的发展,微孔板技术有了长足的进步,但其发展面临着一些不易解决的困难,主要有:微量液体极易蒸发,不适于那些不能用二甲亚砜(DMSO)作溶剂的筛选方法以及受限于当今还不够完善的微量液体分配技术。

微阵列技术是将微孔板技术进一步微型化。最近,哈佛大学的研究人员开发了化合物微阵列芯片,主要用于筛选能与特定蛋白质特异性结合的化合物。他们将玻片表面进行化学处理,使其衍生化产生活性基团,然后将溶于有机溶剂中的化合物用机械手点在经处理的玻片表面,化合物与玻片表面的活性基团反应而被固定于玻片表面,这样就将不同的化合物排布成微阵列,固定在玻片表面,制成化合物微阵列芯片。随后将感兴趣的蛋白质进行荧光标记,然后与微阵列芯片上的化合物反应,经清洗后,再进行荧光检测就可以筛选到能与这种蛋白质特异性结合的化合物。他们用化合物微阵列芯片进行了原理性实验,其结果表明,使用这种化合物微阵列芯片可以平行、快速地进行大规模的化合物与蛋白质的结合筛选。他们最先是将玻璃片表面进行马来酰亚胺衍生化处理,后来采用亚硫酰氯处理,都获得了成功。他们还尝试了使用这种化合物微阵列芯片进行大规模对映异构体的分型检测。加利福尼亚大学Davis分校的科学家们采用类似的方法也制造了一种化合物微阵列芯片。他们对玻璃载玻片表面进行氨基化处理,在氨基玻片上进行乙醛酰衍生化,然后将带有连接臂的配体分子点在修饰过的玻片表面。在进行化学连接反应之后,这种固定了不同小分子配体的微阵列芯片被用来进行了3种生物学检测:蛋白质结合检测、功能磷酸化检测和活细胞粘附检测。实验结果证实了化合物微阵列芯片可以帮助我们对由组合合成方法获得的大量化合物进行快速的功能分析和筛选。化合物微阵列芯片技术与基于微珠体的固相组合合成技术相结合为高通量药物筛选带来了一条新的途径,将对高通量药物筛选技术的发展产生积极的影响。

最近,出现了一种被称为芯片膜片钳(patch - on - a chip)的新技术。在这种膜片钳芯片上加工有检测电信号的点阵,点阵中的每一个点是一个电信号记录单元,同时每个点底部与负极相通,可以吸附细胞。这样膜片钳芯片上的每一个点就可以实现传统膜片钳技术的功能。膜片钳芯片具有操作简单、快速和可实现高通量等优点,可以用于电生理研究和高通量药物筛选。

(二)微流体芯片技术

微流体装置的发展已广泛用于生化及细胞的分析。鉴于这项技术在超高通量筛选中的巨大应用前景,吸引了众多学术界和工业界的实验室对该项技术的研究与开发。借用半导体工业中所用的光刻技术将内径为10~100 μm的通道加工在玻璃或硅片中,利用电动泵和流体的压力来控制皮、纳升级液体的流动。该技术可减少几个数量级的试剂消耗量,并能提高数据质量。它所采用的平行样品处理程序可以获得更高的筛选通量。多种类型的筛选分析方法在微流体芯片上操作的可行性(包括结合分析、酶分析和细胞分析)已得到实验论证。

Jiang等制造了2种可以用于高通量检测食物污染物和进行药物筛选的微流体装置。他们以硅为模板,采用模压和毛细管成形技术在聚酯和多聚硅氧烷二甲酯中加工激流体网络。在第1种装置的微通道中夹着一块PVDF(polyvinylidene fluoride)膜可用来结合、浓缩目标化合

物，然后再直接通过质谱方法鉴定目标化合物。第 2 种装置被用来超滤分离苯巴比妥和苯巴比妥与抗体的复合物。整个过程包括上样、清洗和解离都是连续进行的。这些微流体装置不仅可以显著降低系统死腔体积和所需样品量，而且与先前报道的方法相比灵敏度至少可提高 1～2 个数量级。

Capliper Technologies 公司设计的芯片装置，将从微孔板中取样与随后进行的酶抑制剂筛选结合在一起。目前，该芯片装置已制作完成，并通过了测试。测试的化合物通过与芯片相连的硅毛细管注射入芯片中，与酶及荧光酶底物混合以使它们在主反应通道内发生相互作用。酶与荧光底物反应产生基线荧光信号。如果测试化合物抑制酶的活性，信号将会暂时性的降低，通过检测荧光信号的强度就可以测定化合物抑制酶的活性。这种装置每次进样量 1～10nL，进样同期 10～30 s。利用这个系统，每块芯片可对几千种化合物进行筛选。目前对该系统进行改进的努力主要集中在提高系统的耐用性和将不同类型的筛选方法移植到这个平台上来。最新设计的微流体芯片将液体分发、化合物稀释与筛选分析三者集成起来，由于减少了被视为瓶颈的芯片使用前的准备步骤，使得该系统更加方便实用。

芯片技术具有高通量、大规模、平行性等特点可以进行新药的筛选，尤其对我国传统的中药有效成分进行筛选。如何分离和鉴定药物的有效成分是目前中药产业和传统的西药开发遇到的重大障碍，基因芯片技术是解决这一障碍的有效手段，它能够大规模地筛选、通用性强，能够从基因水平解释药物的作用机理，即可以利用基因芯片分析用药前后机体的不同组织、器官基因表达的差异。如果再用 mRNA 构建 cDNA 表达文库，然后用得到的肽库制作肽芯片，则可以从众多的药物成分中筛选到起作用的部分物质。或者，利用 RNA、单链 DNA 有很大的柔性，能形成复杂的空间结构，更有利于靶分子相结合，可将核酸库中的 RNA 或单链 DNA 固定在芯片上，然后与靶蛋白孵育，形成蛋白质－RNA 或蛋白质－DNA 复合物，可以筛选特异的药物蛋白或核酸，因此芯片技术和 RNA 库的结合在药物筛选中将得到广泛应用。在寻找 HIV 药物中，Jellis 等用组合化学合成及 DNA 芯片技术筛选了 654 536 种硫代磷酸八聚核苷酸，并从中确定了具有 XXG4XX 样结构的抑制物，实验表明，这种筛选物对 HIV 感染细胞有明显阻断作用。生物芯片技术使得药物筛选，靶基因鉴别和新药测试的速度大大提高，成本大大降低。

基因芯片对于药物靶标的发现、多靶位同步高通量药物筛选、药物作用的分子机制、药物活性及毒性评价方面都有其他方法无可比拟的优越性，能够从基因水平解释药物的作用机制，可以用基因芯片分析用药前后机体的不同组织、器官基因表达的差异，国外几乎所有的主要制药公司都不同程度地采用了基因芯片技术来寻找药物靶标，检查药物的毒性或副作用。

四、生物芯片在药物基因组学中的应用

药物基因组学（Pharmacogenomics）是在基因组学的基础上研究不同个体对药物反应的差异以便对不同的基因型患者实行个体化给药，从而将药物的药效充分发挥面使不良反应减少到最小。

将 DNA 芯片技术应用于药物基因组学，一方面可以加速药物基因组学的发展，主要是利用 DNA 芯片进行基因功能及其多态性的研究，以确认与药物效应及药物吸收、代谢、排泄等相关的基因，并查明这些基因的多态性；另一方面 DNA 芯片利用药物基因组学的研究成果，根据基因型将人分群，以实现药物基因组学研究的目的和价值。

临床上，同一种药物对不同患者的疗效与副作用可能不同；并且，不同患者对相同剂量药物的疗效或副作用方面也会有很大差异。这主要是由于病人遗传学上存在差异，如药物应答

基因，导致对药物产生不同的反应。如果利用基因芯片技术对患者先进行诊断，再开处方，就可对病人实施个体优化治疗。

药物基因组学的研究依赖于人类基因组中大量基因变异的确定。近年来研究较多的是单核苷酸多态性（single nucleotide polymorphisms，SNPs）。SNPs 与许多药物效应的个体差异有关，研究它们之间的关系需要定位大量的 SNPs。DNA 芯片可以自动、快速地检测那些可以影响药物效应的基因（如药物代谢酶、药物作用靶和致病因子的编码基因）和决定个人对药物毒性敏感性的基因。Evans 和 Relling 就设想了一种急性原淋巴细胞白血病药物基因组芯片，这种芯片上包括了所有可能影响急性原淋巴细胞白血病病人对化学治疗反应的基因，借助这种芯片可以根据病人的基因型对病人分群，帮助医生为每个病人选择合适的治疗药物和药物剂量。相信在不久的将来，药品说明书上的适用症和禁忌症都会改为适用基因型和禁忌基因型，使得药品更加针对不同个体的不同疾病，达到疗效更佳、副作用更小的目的。

传统的中医药讲究的是整体治疗观，在中药基因组学和中药化学组学中，多成分、多靶点治疗是基点，关键就是分析出成分谱与活性谱，实现精确筛选和配置，以“复方”的形式作用于致病基因组。

五、基因芯片技术在疾病耐药性检测中的应用

详见本章第六节。

六、生物芯片对中药物种的鉴定

中药在现代药物研究中日益受到重视，但中药种类多，其药效受产地、气候、收获季节等多种因素的影响，并且，还存在假冒伪劣中药材的情况，如何鉴别药材的种类和真伪，用现有的方法还难以解决。近年来分子生物学的一些方法开始用于分类学研究，其原理是基于任何一个生物个体，不受发育阶段影响，其不同组织器官部位以至细胞都具备相同的遗传信息。这些方法中以基因芯片技术最有潜力。目前，已有学者开始对地道与非地道药材如贝母、金银花等的 DNA 序列进行研究，分离特异核苷酸序列作为探针并制备 DNA 芯片，进行鉴定。

（赵云斌　张德太）

第八节　生物芯片技术在法医学中的应用

法医学的任务是利用医学和其他自然学科的理论与技术，研究并解决司法实践中有关医学问题的一门医学学科。法医学鉴定为刑事案件侦察提供线索，为审判提供证据，达到揭露犯罪，洗昭冤屈，化解矛盾，公平执法的目的。法医学是法律与医学结合的产物，涉及医学及多种基础学科，法医学的发展与医学密切相关，医学与其他自然科学的迅速发展不断为法医学提供新的理论与技术。20 世纪 90 年代，生物芯片技术研究日渐深入，在医学研究的应用前景极其广阔，法医学生物芯片的应用研究也及时地开展起来。目前进行的探索性、前瞻性研究，已经有了初步阶段性成果，学术界普遍认为这一高、新技术将会使法医学鉴定再上一新的台阶。

一、生物芯片在法医物证学中的应用

法医物证学的主要任务是通过人类遗传标记的检测与分析进行个体识别与亲子鉴定。个体识别又称同一性鉴定。现代刑事犯罪理论认为：接触就会遗留痕迹。刑事犯罪是一个客观过程，从刑事案件现场收集到罪犯遗留的生物学检材，例如血痕、精斑、唾液、毛发等均是证实

犯罪的物证。通过对现场材料和犯罪嫌疑人的遗传标记的检测与比对，可以达到排除或认定罪犯的目的。法医学亲子鉴定是通过对假设父亲，母亲和孩子的遗传标记的检测，分析亲代与子代间是否符合孟德尔遗传理论，判断是否亲生关系。

（一）遗传标记应用现状

法医学应用的传统遗传标记是血型，进入 20 世纪 80 年代，遗传标记的分析进入 DNA 分子水平。1985 年，法医 DNA 的限制性片段长度多态性（restriction fragment length polymorphism, RFLP）检测方法建立，检测手段主要是 Southern 杂交。采用人类小卫星 DNA 核心序列作为探针，在低强度杂交条件下，同时与多个可变数目串联重复序列（variable number of tandem repeats, VNTR）基因座形成杂交产物。杂交形成的 RFLP 图谱由十数条不同的片段长度组成，形似商品上标记的条码。图谱具有高度个体特异性，不同个体间片段数不同、片段长度不同，无关个体间的平均匹配概率达到 $10^{-12} \sim 10^{-19}$。DNA－RFLP 检测使法医学第一次能够认定同一性，因此小卫星 DNA－RFLP 分析被誉为 DNA 指纹（DNA fingerprint）。如果控制杂交条件，在较高的杂交温度、较低的离子强度时，小卫星探针只能与一个 VNTR 基因座的靶片段杂交。形成的 RFLP 图谱只是一个座位的等位基因型别，杂合子二条带，纯合子一条带。因小卫星 VNTR 基因座具有极高的多态性，单座位 RFLP 图谱也具有高度个体特异性，联合检测 4～6 个单座位 VNTR，同样可以达到 DNA 指纹的识别能力。为与多基因座 VNTR－RFLP 图谱区别，单基因座 VNTR－RFLP 图谱被称为 DNA 纹印（DNA profile）。DNA 指纹技术克服了血型标记检测的许多局限性，在案件调查中发挥了重要作用。多年的实际应用发现，DNA－RFLP 检测方法烦琐，检测时间长，检测灵敏度低，有关结论解释的理论基础尚待完善，在实际案件材料的检测应用时受到限制。

第二代 DNA 遗传标记是应用 PCR 技术扩增人类基因组中 VNTR 序列。在卫星 DNA 中，VNTR 大约占一半，具有串联重复的特征性序列，以相对恒定的序列片段作为重复单位，在某个基因座位串联重复。不同个体、不同等位基因间表现为重复的次数不同，构成复杂的片段长度多态性。直接应用 PCR 扩增 VNTR 座位的等位基因，然后电泳分析基因型，避免了 RFLP 过程中的限制酶切，Southern 印迹转移，探针标记与杂交等烦琐的操作，极大地简化操作程序。目前，法医 DNA 分型检测常规采用微卫星 DNA－VNTR 序列遗传标记，这类卫星序列的重复单位长 2～6 bp，重复次数由数次至数十次，等位基因片段长度均在 400 bp 以下，故称短串联重复序列（short tandem repeat, STR）。PCR－STR 分型具有高灵敏度、高多态性的优势，适用于腐败、陈旧检材的法医学鉴定。目前已经开发出 PCR－STR 分型的自动检测系统，以荧光染料标记 PCR 引物作多 STR 基因座复合扩增，扩增产物经毛细管电泳或聚丙烯酰胺平板凝胶电泳，利用激光诱导方式自动进行 STR 基因型分析，结果记录和数据贮存。荧光标记 PCR－STR 分型已成为法医物证鉴定的标准技术。

1997 年，单核苷酸多态性（single nucleotide polymorphism, SNP）标记策略的提出，在整个生物学界引起一个 DNA 多态性研究的新热潮。随着人类基因组计划研究的深入，学者们对基因组的碱基变异有了更加深入的认识。在物种长期进化中，基因组 DNA 序列不断地发生变异，其中一部分被保留下来，导致不同种属、不同群体和不同个体间的差异，形成多态性。SNP 是指基因组内某特定的核苷酸位置上存在有 2 种不同的碱基成分，按照多态性的概念，其中少的一种碱基在群体中的比例应不少于 1%。与 STR 相比，SNP 在基因组中有更广泛地分布，如果按照 0.5%～10% 的碱基变异率计算，人类基因座中至少每 1 kb 就有一个 SNP，总 SNP 数至少有 300 万，分布密度远多于其他 DNA 遗传标记。由于选择压力的原因，包括 SNP 在内的碱基

序列变异在非编码区的分布要远多于编码区，据估计人类编码区中的 SNP 大约只有 20 万个。从生物遗传和变异的本质分析，SNP 可能是一种最能反映群体遗传结构和个体间的遗传差异的标记系统。

SNP 是单个核苷酸变异形成的 DNA 序列多态性，理论上一个核苷酸位置可以有 4 种碱基变异形式，但实际上 SNP 多表现为双等位基因，即二态性遗传变异。碱基变异包括单碱基的转换、颠换以及单个碱基的插入与缺失等，其中转换与颠换的比例为 2∶1，有 25% 的 SNP 发生在 CpG 位点。虽然 SNP 的多态性程度远不如小卫星或微卫星，但是 SNP 数量巨大，基因组中分布频密，从整体而言，SNP 的多态性实际上要大得多。由于每个 SNP 位点均为二等位基因，在 SNP 检测时能通过简单的'+/-'方式进行表型分析，无须测定基因的片段长度，检测结果易于自动化。SNP 将成为法医物证鉴定应用的第三代遗传标记系统。SNP 大约包含了人类基因组中已知多态性的 80%，是最常见的遗传变异形式，在人类基因组学，医学遗传学，疾病相关性研究，药物敏感性相关基因研究等以及法医学等方面将发挥重大的作用。

法医学所关注的 DNA 遗传标记除常染色体外，有些特殊遗传的标记系统在法医学鉴定中常有应用，例如线粒体 DNA 和 Y 染色体上的某些标记系统。

线粒体 DNA（mtDNA）是唯一的核外基因组，每个线粒体平均含有 10 个 mtDNA，每一个体细胞大约有 10^5 个 mtDNA。mtDNA 分子小，仅 16 569 bp，呈环状，分子裸露，编码 tRNA，rRNA 和部分与氧化磷酸化有关的蛋白质。序列分析发现 mtDNA 大约有 1/3 的碱基替换突变率较高，因为：①mtDNA 分子裸露，易受各类致损因素的影响；②mtDNA 的 DNA 聚合酶复制精确度及编辑能力较低；③缺乏较完整的 DNA 修复系统；④承受的选择压力较低，碱基变异容易固定。mtDNA 的进化大约比核基因组快 10 倍，突变率约高 10 倍。mtDNA 碱基变异主要集中在非编码区，称为控制区，约 1 100 bp。在控制区碱基变异不是随机的，相对集中在二个高变区，HVR－Ⅰ和 HVR－Ⅱ。mtDNA 序列多态性的法医应用有三个特征，一是灵敏度高，原因是 mtDNA 拷贝数远多于核基因组 DNA，体细胞中核基因组只有二个，而 mtDNA 的拷贝数有数千或数万。高度腐败，陈旧、超微量材料，甚至无核组织如毛发，惟有提取和检测 mtDNA 型可能成功。二是 mtDNA 序列具有高度多态性，HVR－Ⅰ和 HVR－Ⅱ中存在大量的点突变，无关个体间序列匹配概率的机会很小，是个体识别鉴定有价值的标记系统。mtDNA 遗传标记的第三个特征是具有母系遗传特征，在确定母子亲生关系的亲子鉴定有特殊作用，在重建母系家族关系，研究人类进化，人种的地域分布、演变及大规模的迁移特征有实用价值。mtDNA 的多态性检测主要采用序列测定方式，目前已经报告了多个人种、地区 mtDNA 多态性的群体资料。限于技术及设备的要求较高，mtDNA 多态性检测在法医学鉴定应用中受到限制。由于 mtDNA 基因组的点突变主要以二碱基变异的 SNP 形式出现，应用生物芯片技术检测将是一个有效的手段，近期已有应用芯片技术检测 mtDNA 的 SNP 的研究报道。

Y 染色体上的遗传标记，受到法医学和人类进化研究的关注，Y 染色体约 60 Mb 核苷酸，占人类基因组的 2%。在染色体长臂和短臂各有一拟常染色体区，长度分别有 2.6 Mb 和 0.32 Mb。Y 染色体上存在单碱基突变，碱基的插入、缺失和 VNTR 序列变异，构成多态性标记。Y 染色体除了拟常染色体区段外，在减数分裂时大部分不与 X 染色体重组，这些非重组部分多态性标记只能从父亲遗传给儿子，属于一种单体父系遗传方式。对于男性嫌疑人的认定，男性个体间亲权关系鉴定案件的调查以及重建父系家族关系时有特殊的意义。Y－STR 基因型在不同群体的构成有差异，在不同个体间有差异，是目前法医学鉴定常用的标记系统。近年来，随着对 Y 染色体序列研究的深入，Y 染色体上的 SNP 的分布与分型在法医学的应用研究已经

成为一个热点。

(二)遗传标记系统多态性评估

为达到较高的识别能力,法医学选择遗传标记时通常首选高多态性标记系统。多态性程度高,识别能力高,法医学鉴定中的实用价值就高。评估多态性高低常用遗传学指标有杂合度(heterozygosity,H),法医学应用评估指标有个体识别能力(discrimination power,Dp)和非父排除率(probability of exclusion,PE)。三个评估指标均可通过相应的计算公式计算。影响多态性高低主要有两个因素,一是基因座上的等位基因数,二是群体中等位基因频率分布。一般情况下,遗传标记系统的等位基因数越多、等位基因频率分布越平均,多态性程度越高。评估多态性的三个指标主要通过群体调查,观察等位基因数和计算基因频率,然后可以得到该基因座的群体遗传参数。

法医学个体识别鉴定和亲子鉴定结论都是双向性的,要么排除同一性或亲生关系,要么是认定同一性或亲生关系。选择高多态性遗传标记的目的是要求达到遗传表型具有高度的个体特异性,鉴定结论能够是排除是绝对的,认定也是绝对的。虽然从统计学理论分析,不可能达到这一终极目标,但当某一种基因表型是一种极其罕见的小概率事件时,接近的绝对排除或认定结论是有可能的。例如某现场的精斑与嫌疑人的检材经过 12 个 STR 基因座检测,所有的基因型均相同,计算累计匹配概率为 10^{-13},意思是在一千亿无关个体中,此遗传表型是唯一的,此时认定嫌疑人是罪犯应毫无疑问。当一个嫌疑父亲在检测 13 个 STR 基因座后,仍不排除他是有争议孩子的亲生父亲,计算似然率超过 2 000,父子关系相对机会超过 99.95%,认定亲生关系也应无疑义。累积平均匹配概率达 10^{-13},累计非父排除率达 0.999 5 是目前国际上通行的统计学衡量指标。目前常用的 STR 基因座多态性程度比较低,二态座位 SNP 多态性更低,唯一的办法是同步检测多个基因座,这是法医物证学近年来应用研究的重点。

要达到一个公认的规范的鉴别能力标准,至少要检测多少个遗传标记系统,是法医学应用所关注的问题。法医学应用选择的 STR 基因座,等位基因数一般在 10 ~ 20 个,杂合度 0.7 左右,多态性程度较低,复合扩增 13 ~ 16 个基因座一般能够达到鉴定要求。SNP 座位为二等位基因结构,同步检测多少个 SNP 才能够达到专业衡量指标,从多态性理论推算或者实际案件检测的角度分析,SNP 基因座检测的基本数目应予确定。Delahunty 等(1996)首次应用寡核苷酸连接试验(oligonucleotide ligation assay, OLA)对多 SNP 基因座同步分析进行了尝试。针对每一个 SNP,设计有检测探针和示踪探针,均含约 20 核苷酸,分别用生物素和同位素 ^{32}P 标记。OLA 反应中,先行扩增样品 DNA,两种探针均与样品扩增产物杂交。检测的特异性的关键在检测探针的 3′ – 末端的碱基是否与靶片段匹配,如果是匹配关系,反应体系中的 T4 连接酶则将检测探针与示踪探针连接起来,形成稳定的"检测探针/靶 DNA 扩增产物/示踪探针"三明治状杂交产物,示踪分子 ^{32}P 显示阳性结果。作者选择了 11 条染色体上的 19 个 SNP 基因座:AT3,PROS1, BCHE, ARSB, CG26, VB14, VB6.9, TYR, IGF2, VWF, COL2A1, F7, CA3, HT3, VA28, VA23, BCL2, LDLR 和 PRNP。每位点的二个等位基因,相应设计的 2 个检测探针均用生物素标记,示踪探针采用 dUTP – 毛地黄毒苷标记,显色反应利用抗毛地黄毒苷抗体 – ELISA 系统。阳性为深红色,阴性为无色,以微反应板色谱计测定 490nm 处的吸收光谱值,大于 0.15 为阳性,低于 0.15 为阴性。检测结果转化为数字:基因 1 阳性、基因 2 阴性基因型的个体为 1 型;基因 1 阴性、基因 2 阳性者为 2 型;基因 1、2 均为阳性杂合子个体为 3 型;基因 1、2 均为阴性为 4 型,其中 4 型代表 PCR 扩增失败,此型占总个体数不到 1%。作者根据群体调查确定 19

个 SNP 座位的二基因频率比值平均为 62:38,计算出 19 座位的累计平均匹配概率为 4.8×19^{-9}。在基本方法确定后,所有的加样,洗涤杂交过程均可利用机械手操作平台完成,作者对 PCR – OLA 法同步检测多个 SNP 多态性问题作了有益的探讨。

Brenner(2001)认为,达到法医鉴定需要的鉴别能力,至少需要多少个标记系统,受多个因素影响。对于个体识别鉴定而言,一个单基因座 VNTR – RFLP 检测的多态性,相当于 2 个 STR 座位,或 6 个 SNP;而对于亲子鉴定,一个 VNTR – RFLP 系统相当于 3 个 STR 或者 12 个 SNP,如果用于典型的混合斑个人识别鉴定,这个比率则为 1:4:27。目前学术界基本能够接受的意见是,个体识别鉴定需要 50 个常染色体 SNP,亲子鉴定要 100 个 SNP 基因座,才能够达到目前 PCR – STR 分析系统的鉴别能力。发现与确定 SNP 基因座目前不是难题,现在公共数据库中已经有 100 万个以上的 SNP(http://snp.cshl.org/)。利用生物芯片技术检测展示的多 SNP 组合,单从多态性程度而言,50 – 100 个 SNP 基因座组合在理论上是能获得可靠的鉴别力,不过,在案件鉴定中的信息可信度还要通过实践的应用评估。

(三)生物芯片在物证鉴定中的应用

生物芯片根据生物化学反应过程分类,分为样品制备芯片,生化反应芯片及各种检测芯片等。检测技术的核心是核酸分子杂交,借鉴大规模集成电路制作的思路,设计不同的寡核苷酸探针,按照预先设计的阵列顺序将众多的探针固定在载体支持物上,然后与标记的样品分子同步进行杂交。或者将不同的样品靶 DNA 分子固定在载体表面上,然后与不同的探针杂交。检测每个探针分子的杂交信号阴性或阳性以及杂交信号的强度可以获得样品分子的数量和序列信息。由于探针的点阵列是高密度的,探针的原位合成技术和照相平板印刷技术结合以及激光共聚焦显微技术的应用,使核酸杂交实现集成化、高通量,使一次性对样品大量序列进行实时、灵敏、准确的检测与分析。DNA 芯片开拓了一个同步检测多个遗传标记系统的新技术,在法医学物证鉴定中将发挥重要的作用,而样品制备芯片和生化反应芯片(PCR 芯片)在法医物证学中的应用研究也有一定的发展。

1. 微量核酸样品的制备。在法医实践中经常遇到样品量稀少的情况,用传统的方法进行核酸的分离和纯化非常困难,它包括了细胞分离、破胞、脱蛋白、提取 DNA、PCR 等多方面的工作。而生物芯片的构造特点决定了在生物芯片上进行样品制备只需要极微量的样品,这在法医 DNA 分析中具有很高的实用价值。宾夕法尼亚大学的研究小组研究的一种微过滤芯片已经能够从血液中得到高收率的人白细胞。为了把不同的癌细胞、细菌或病毒从血样、水样或其他液体样品中分离出来,人们设计加工了介电电泳(dielectrophoresis)芯片,用该芯片现已成功实现了对乳腺癌细胞和正常单核白细胞的分离。芯片中的破胞方法有芯片升温破胞、变压脉冲破胞,以及化学破胞等。在捕获 DNA 方面,Cepheid 公司应用湿法蚀刻和反应离子蚀刻/等离子蚀刻等工艺在硅片上加工出含有 5 000 个高 200 μm,直径 20 μm 的具有细柱式结构的 DNA 萃取芯片,专门用于提取低浓度(低于 105 个目标 DNA)和高浓度(100 ~ 1 000 ng/mL 量级)的 DNA。目前,Cepheid 公司已经开发出功能更强大的样品制备芯片。由于生物芯片样品制备技术和芯片 PCR 技术所需样品量少、反应速度快、无交叉污染,所以,该技术用于法医 DNA 分析将大大提高法医鉴定的速度和鉴定结果的准确性。

2. STR 基因座分型检测。STR 基因座多态性本质是 VNTR 序列,法医学应用的座位大多数是四核苷酸 STR 座位,少数是五核苷酸座位,分型稳定,重复性好,等位基因片段一般在 400 bp 以下。PCR – STR 分型是目前法医学鉴定的主要手段,常规采用多座位复合扩增方法提高单

次检测的信息量，累计多个基因座的鉴别信息，可以达到 10^{-12}的平均匹配概率。STR 座位片段长度等位基因的鉴别需要高分辨的电泳技术，区别数 bp 长度差异的等位基因，技术要求较高。应用生物芯片技术可以避免电泳程序可能出现的误差，应是一项有前途的技术手段。但采用芯片技术检测多个 STR 基因座作过多次尝试，效果均不够理想。在 VNTR 座位，相差一个重复单位就形成一个等位基因，STR 的各等位基因长度间差别可能只有 4bp，重复单位的序列相同或只有个别碱基差异，以常规芯片的被动杂交方法检测时，等位基因探针与靶基因片段常出现滑动杂交(slippage hybridization)，出现假阳性结果，使等位基因和基因型的误判。常用来解决杂交忠实性的办法是提高杂交温度和降低离子强度，人为提高核酸分子杂交条件的强度。但是这种人为控制的杂交条件必须在 DNA 阵列的每一个点上是一致的、均一的，才能使整个芯片上所有点上不同序列的 DNA 分子同时完成杂交。显然，采用被动杂交、又要求杂交的均一性和忠实性，操作难度大，过程烦琐费时，试验一直未获满意的效果。

Radtkey 等(2000)根据 DNA 杂交动力学原理，利用不同碱基堆积能(stacking energy)的差异，设计由电场控制的主动杂交微阵列系统检测 TH01 和 CSF1PO 两 STR 基因座获得成功。该技术的主要优点是利用电场作用，完成探针与靶 DNA 片段主动杂交过程。在人为控制的高强度杂交条件下，避免了芯片寡核苷酸阵列中重复序列位置上的滑动杂交产物形成。以 TH01 座位为例：

靶片段序列：5′ – ACACAGACTCCATGGTG[AATG]$_n$ACGGAAATAAGGGAGGA – 3′

其中 $n = 4 \sim 10$，表示此基因座有 4,5,6,7,8,9 和 10 等 7 个等位基因。

针对 TH01 座位检测，分别设计捕获(capture)寡核苷酸探针分子和示踪(reporter)寡核苷酸分子，分别以生物素(biotin)和荧光染料 Bodipy Texas Red (BTR)标记。

捕获分子：5′ – biotin – TCCTCCCTTATTTCCCT[CATT]$_{4\sim10}$ – 3′

示踪分子：5′ – CATTCACCAT – BTR – 3′

捕获分子的定位，靶序列扩增片段杂交，示踪分子的定位均依靠电场的作用，故微芯片的制作与被动杂交装置不同，每一个阵列点都联系着电极(图 17 – 32)。

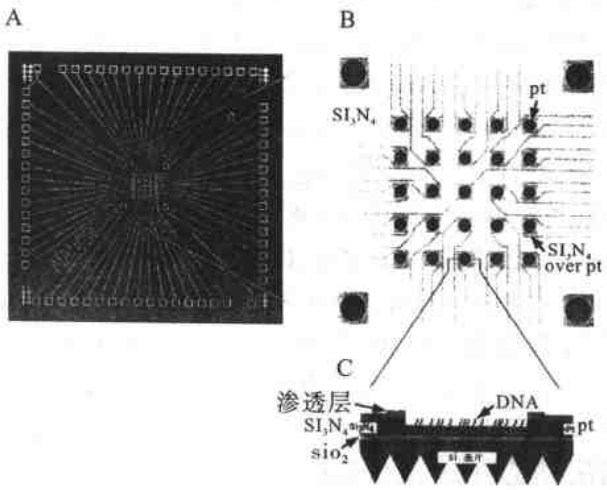


图 17 – 32　微电场 DNA 芯片装置(Ronald)

分型检测基本过程如下：①寡核苷酸捕获分子经过电场作用分别定位在芯片的特定点的衬垫上，采用交流电每衬垫点电流 0.8 μA，以正电 38 μs 和负电 25 μs 转换，共 25 s。②样品 DNA 扩增产物以电杂交方式与捕获分子杂交结合，每衬垫点电流 1.6 μA，以正电 19 μs 和负电 12 μs 转换，共 3 min。为预防捕获分子与靶 DNA 出现滑动杂交，此杂交温度控制在 32～34 ℃。洗涤去除多余的靶 DNA。③ 加入示踪分子溶液，示踪分子与靶 DNA 片段的 5′端杂交结合。此时已形成“捕获分子/靶 DNA 分子/示踪寡核苷酸分子”的三明治复合物，30～40 ℃温度条件下洗涤数次，去除错配的等位基因。④ 阵列荧光图像检测，记录结果。

从图 17－33 表示的基本原理中可以看出，设计的关键是捕获分子中的重复单位数。捕获分子加示踪分子的重复单位数的和小于靶等位基因片段重复单位数时，二者间会出现空缺；重复单位数和大于靶片段时，二者间会出现重叠，均为错配，不能形成稳定的三明治复合结构，30～40 ℃温度条件下洗涤可以消除暂时结合上的不稳定的示踪分子。只有当捕获分子与示踪分子的重复单位总和等于靶等位基因重复单位数时，才能形最成稳定的杂交复合物，显示阳性结果。

图 17－33　分子杂交检测 STR 基因型的寡核苷酸系统(Radtkey)

Radtkey 等建立的电场控制的主动杂交 DNA 芯片方法解决了两个关键技术问题。在电场的控制下，完成捕获分子的定位和靶 DNA 片段的杂交，同时具有浓缩寡核苷酸的作用，提高杂

交效率和检测的灵敏度，50 ng 模板 DNA 就可以完成 STR 分型检测。电杂交明显缩短了检测时间，捕获分子与靶基因片段杂交可以在 5 min 内完成。另一个是针对各不同等位基因设计捕获分子，是准确检测出等位基因的关键因素。在人为控制的较高温度、较低离子强度的杂交条件下，捕获分子与靶分子才能达到完全匹配，碱基"堆积"的杂交复合物最稳定，才出现阳性结果，可以准确地判定样品的基因型。在某些复合 STR 基因座，重复单位长度相同但碱基序列间有微小差变异，或存在插入等位基因（如 TH01 座位的 9.3 基因），也能采用该法准确检测，只要合成相应的捕获分子寡核苷酸就能获得阳性结果。

按照目前法医学鉴定对遗传标记的鉴别能力要求，一般需要同步检测 13～16 个 STR 基因座，以每基因座 12 个等位基因计算，大约 200 个杂交点构成的芯片阵列就可以完成同步检测，这在芯片的制作上不是难题。采用芯片技术分型 STR 基因座避免了等位基因片段长度的测定，简化检测程序，减少人为误差，具有一定的技术优势。电场控制的主动杂交 DNA 芯片技术分析多 STR 基因座，理论上的论证是可行的，但是有关 STR 基因座的选择与组合以及该方法能否适用于各类陈旧、微量、腐败生物学检材的检验鉴定，目前尚在研究中。

3. SNP 基因座分型检测。SNP 在基因座中高密度分布，具有稳定的遗传。因为是二等位基因结构的标记，以单个座位比较而言，SNP 的多态性是最低的。从法医学应用的要求分析，为达到最大的鉴别能力，选择 SNP 可以参考以下几个原则：①等位基因频率在 0.2～0.8 之间。基因组中，并非所有的 SNP 都适合作遗传标记，也并非只有 0.5/0.5 对等频率分布的 SNP 识别能力最高。有作者研究，利用基因频率为 0.2～0.8 的 SNP 标记构建高密度的基因遗传图谱，效果与 0.5/0.5 频率分布的 SNP 无明显差异。②SNP 座位间距离较近（1～2 cM），因而可以同步检测更多的基因座，提高单次检测的信息量。③突变率低，SNP 座位的遗传稳定性较好，相对于 STR 具有一定的优势。国际法医遗传学学会 DNA 委员会对人选的 STR 座位的一个基本要求是突变率低于 0.002，在筛选 SNP 座位时，至少要参照这一标准。④法医学应用研究评估，包括个体同一性，不同检材基质，不同保存条件、时间，不同的物理、化学及生物学因素对 SNP 分型的影响等问题需要研究。还有 SNP 的种属特异性，检测灵敏度等实际检测中常遇到的问题等。

SNP 遗传标记属于 DNA 序列多态性，所有的序列多态性检测方法都可以利用。传统的限制性片段长度多态性技术（RFLP），单链构象多态性（SSCP），异源双链体测定技术如变性梯度凝胶电泳（DGGE）和温度梯度凝胶电泳（TGGE），等位基因特异性 PCR（AS－PCR），DNA 序列测定等。但这些技术均依靠电泳技术，难以进行大规模、自动化、多座位同步检测。针对 SNP 的高通量检测，近年来，已研究有多种新方法。其中生物芯片技术最具有吸引力，在样品 DNA 制备，探针的合成与固定，分子的标记、数据的读取与分析都取得长足的进步。芯片技术的基本原理是 DNA 分子杂交，在固相载体表面上直接合成寡核苷酸探针，或采用高精密度机械手将合成好的寡核苷酸探针直接点样在载体表面上，构成密集的分子阵列。再与靶 DNA 扩增产物杂交。根据激光激发杂交信号的不同波长区分阳性与阴性结果，由激光共聚焦显微镜检测、收集信号，计算机处理数据。目前，法医物证学研究的热点是应用生物芯片技术检测 SNP，重点在基因座的选择，群体资料调查，检测方法的建立。虽然尚在研究阶段，虽然某些技术细节仍存在困难，但芯片技术的巨大潜力不容置疑。

SNP 座位的等位基因间相差仅一个碱基，探针与靶 DNA 片段的杂交常会出现错配杂交产物，导致假阳性结果。Gill 等（2001）设计一种等位基因特异性通用示踪引物，建立起多 SNP 座位复合扩增检测分型系统方法。作者设计了两套引物，一套是起始引物，先行扩增含有 SNP

座位的靶片段。另一套为检测引物，是针对二等位基因 SNP 基因座设计的。每一个 SNP 基因座要有两个正向(F)引物和一个反向(R)引物，这些座位特异性引物的 3′端含有 16～27 碱基，与靶 SNP 相应区域互补，引物的 5′端含有 20 碱基，与靶序列不互补，称通用序列。二个正向引物的 5′端通用序列不相同，3′末端碱基分别与 SNP 多态性碱基对应，也不相同，而反向引物中通用序列是一致的。在 PCR 退火阶段，只有 3′末端碱基与 SNP 多态碱基互补时，引物才能继续延伸形成扩增产物，否则没有产物。检测结果，杂合子有两个扩增产物，纯合子只有一个扩增产物。作者的这种设计使决定等位基因特异性有两种因素，一是 5′端含 20 碱基的通用序列，其次是引物 3′末端碱基，分别是对应关系。通用引物不能直接从模板 DNA 扩增出 PCR 产物，但是能够从起始引物的扩增产物中延伸扩增，所以在第二次扩增中，决定等位基因特异性主要是靠由 20 碱基组成的通用引物。因此，极大地提高了检测特异性，最大限度地减少了在直接扩增中因引物 3′端一个碱基错配导致假阳性的可能性。两个通用序列和一个反向通用序列可以从起始扩增产物中分别扩增出 SNP 的 2 个等位基因。针对 SNP 两个等位基因的通用序列分别用 FAM 和 JOE 荧光标记，扩增产物既可以用 Prism377 凝胶自动测序仪器分析，也可以采用生物芯片技术检测。作者已经测试了 80 多个 SNP 基因座，调查了 3 个群体的频率资料，取得满意的效果。目前检测 10～13 个 SNP 的同步检测系统产品正在研制并开始常规生产，作者有信心设计出更多 SNP 座位同步检测系统产品，供法医学鉴定应用。作者还强调，该方法设计的 PCR 体系，扩增产物长度不到 100 bp，相信更适用于腐败、陈旧和微量的法医生物学检材的分析鉴定。

近年来 SNP 的检测方法的研究有了多种探索，Divne 等(2001)利用液相微测序技术建立起多 SNP 座位分析的方法。先利用机械手将各探针 DNA 片段点样在一张硅烷化处理的载玻片上，经过甲醛处理，形成微阵列。作者选择 C/T 变异的 SNP 座位，以一种荧光染料 TAMRA 标记 ddCTP 和 ddTTP，分两个 PCR 反应体系进行。PCR 过程中，引物链以双脱氧核苷酸末端形式终止在多态碱基处。然后与微阵列中的探针杂交，形成的杂交产物带有荧光标记物为阳性，在激光共聚焦显微镜下观察并记录结果。为了获得高灵敏度的效果，作者设计的 PCR 系统扩增产物很短，极有利于微量、陈旧和高度腐败检材的分析。利用这一检测系统，作者一次分析了 13 个核基因组 SNP 和 21 个 mtDNA 的 SNP 基因座，样品包括部分实际案件材料及已知基因型的对照样品，均获较理想的结果。研究结果说明，方法建立后，可以增加被检测的 SNP 数，为法医鉴定提供一个灵敏、可信、快速的高鉴别能力的多 SNP 分型系统。

Lareu 等(2001)建立的 DNA 芯片技术分型 SNP 是利用 Affymetrix 417 微阵列和 Affymetrix 418 微阵列自动扫描仪。作者检测了 100 个常染色体 SNP 和 32 个 Y－SNP 基因座，先将探针在一张载玻片上点样形成阵列，靶 SNP 的两等位基因扩增引物分别标记 Cy3 和 Cy5，然后进行杂交。结果检测和型别判定由微阵列扫描仪完成。

生物芯片对 mtDNA 的分型技术也有一定的发展。Reynolds 等用 17 条 SSO 探针对 689 个人的 HVII 区的多态性进行了分析，还对五例案件的头发和血痕进行了分析，认为该分析法可用于个体识别。Gabriel 等用 27 条探针的微阵列对 105 个个体的 HVI/HVII 内的 6 个区进行了分型，建立了一个 mtDNA 单倍型数据库，并通过分析毛干及陈旧骨骼等实际案例检材，证明该分析方法可准确快速地用于法医鉴定中该类检材的分析。另外，Affymetrix 的 Mark Chee 等还构建了 13 500 条探针的高密度芯片，用以对整个 mtDNA 进行分析。

作者认为，在案件材料鉴定时，对常染色体 SNP，mtDNA－SNP 和 Y－SNP 的分析，微阵列的设计应有区别。对杂交条件的控制，分型原则，检测灵敏度等应用关键问题作了探讨。

法医物证学学科发展一直在不断吸收其他相关学科的理论和技术，DNA 芯片技术的高信息量的优势具有吸引力。在医学诊断领域，对 SNP 进行了大量专题研究，但法庭科学的要求与临床医学不同，应用研究尚在起始阶段。法医学应用面临的挑战是：①基因座的选择，通过群体调查确定基因频率，从多态性程度高低考虑，应选择频率在 0.2～0.8 范围内的 SNP。②建立同步检测 50 个以上的复合检测系统，评估多 SNP 检测操作平台，大约包括 10～50 个 PCR 扩增反应，要求平衡、稳定地扩增，并有足量的产物。③检测灵敏度不低于 10 ng 模板 DNA。④质量控制体系，防止样品间交叉污染及其外来污染。作为一种高新技术，法医生物芯片检测 SNP 的研究将会继续，是机遇，也是挑战。Gill(2001)在第 19 届国际法医遗传学会上对此发表了他的意见，认为 SNP 目前只有检测 DNA 序列变异的价值，以 SNP 代替目前常规应用的 STR 遗传标记目前只是一种推测。现行的国家级的罪犯 DNA 数据库都是以 STR 为基础的，由于 SNP 与 STR 不能够兼容，如果以 SNP 替代 STR 将会引来许多难以克服的管理上的问题。Gill 认为，在 Y－SNP 和 mtDNA－SNP 分型将在法医个体识别中占有一席之地，预期将是目前方法的一种补充，而不能替代现有的方法。也有学者认为目前的 STR 分型将会维持到 2010 年，取而代之的是芯片技术，这是一种较为乐观的估计。法医学鉴定有其自身的特点与要求，从目前的应用研究到成熟的标准化产品面市，尚需时日。

二、生物芯片技术在法医病理学中的应用

(一)生物芯片技术与病理学进展

对疾病组织的表型特征进行形态学分析仍是目前诊断病理学的常用方法。除此以外，许多病理学实验室还通过免疫组织化学、定量分析或电子显微镜获取用于诊断的额外信息。近年来，由于分子生物学的飞速发展，许多临床医生和基础研究人员将目光集中在疾病的遗传改变和基因表达分析上，因为这是疾病生物学特征及相应临床表现的基础。现代医学的核心是揭示疾病的发生、发展机理，其中基因或基因组水平上的分子病理学研究则是最新、最深层次的病理学。

人类所有的疾病都直接或间接地与基因有关，因此监测疾病的全基因组表达和遗传改变具有重要意义。然而，现有的实验室技术不适合于在 mRNA 水平(如 Northern 印迹、RT－PCR 和 ISH)或蛋白质水平(如 Western 印迹和 IHC)检测大量基因的表达。同样，对遗传改变的检测，目前的技术要么分辨率高但检测量有限(如用突变分析检测单碱基改变、Southern 印迹杂交、直接测序和 ISH)，要么检测量大但分辨率有限(如用比较基因组杂交技术分析染色体水平的改变或培养细胞的核型分析)。同时，所有这些技术相对复杂且需要特殊的设备和专业技能，这在许多实验室还无法满足。此外，这些技术劳动强度及耗费均很大。所有这些不利因素限制了其在诊断病理学中的广泛应用。

人类"生命登月计划"——人类基因组计划(HGP)的实施和完成使得人类物理和遗传图的构建及人类疾病基因的鉴定取得了突破性进展，同时这一生命科学的里程碑昭示着人类的基因组研究将步入后基因组时代。利用基因组序列信息发展起来的一些新技术引发了目前生物医学研究领域的革命。在这些新技术中，以 DNA 芯片为代表的生物芯片(biochip)技术克服了传统技术的诸多缺点，成为当前生物医学研究领域最有生命力、最前沿的新技术。

生物芯片技术是新近发展起来的一门新兴技术，它可以将成千上万乃至几十万个与生命相关的信息集成在一小的载体(玻片、硅片或尼龙膜)上，对基因、抗原活体细胞和组织等进行测试分析。近年来，生物芯片技术得到了迅猛发展，目前已有多种类型的芯片出现，如 DNA 芯

片、蛋白芯片、组织芯片等，这些芯片技术具有低劳动强度、高分辨率、高通量的优势，目前已广泛应用于生物医学研究的各领域。同样，在病理学方面，生物芯片技术的出现极大的推动了分子病理学的发展。

人类基因组草图绘制完成后，生物医学研究的重点将转移到了解不同基因的功能上来，即进入功能基因组时代。基因表达是一个高度复杂并相互作用的过程，包括众多基因在时间和空间上快速、动态改变。人类基因组中约有 3.5 万种基因，每种人类细胞一次仅表达一小部分，这种转录组的构成决定了每种细胞类型的独特功能。疾病会引起组织中基因的表达发生变化，对大量基因的表达进行分析研究将有利于阐明不同基因间的相互作用及其与疾病的关系。生物芯片技术可以平行方式直接检测大量基因的表达，从而获取不同组织中个体基因表达的信息。用生物芯片技术分析大量基因的表达可进行基因簇聚、疾病分类、分子诊断及疾病相关基因的筛选。染色体和基因删除、扩增和非整倍性可导致许多肿瘤组织内基因数目异常，对每个细胞内的基因拷贝数进行定量可以帮助我们了解疾病过程。微生物的检测和分类的主要目的是通过探针 DNA 非竞争性和已知来源的细菌或病毒 DNA 或 RNA 的阵列杂交来检测这些微生物特异核酸序列的存在。这方面的典型例子是人类免疫缺陷病毒(HIV)芯片。此外，通过突变和/或多态性研究可进行疾病的易感性分析，根据生物芯片提供的基因信息进行药物筛选和治疗方案的选择等。

组织芯片，又称为组织微阵列(tissue microarray，TMA)，是 1998 年由 Kononen 等在 cDNA 微阵列的基础上发明的，它是将组织高密度地固定于固相载体上，然后用不同的基因、寡核苷酸、抗体与之进行杂交或反应，以研究目的基因或蛋白在不同组织之间的差异表达情况。

在构建组织微阵列时，首先根据 HE 染色结果确定待检组织的基本情况(如肿瘤的大小、类型、分级等)，再确定预计取样组织的位置，然后用不锈钢针进行穿刺(内径为 0.6 mm)，穿刺针附有支架，以确定 X、Y 轴方向，保证穿刺的准确性。一般取材深度为 2～3 mm，将此柱状物固定于一事先钻好孔的模具上，每个标本间隔 0.1 mm，再以 0.4 μm 的厚度进行切片，将此薄片粘附于尼龙膜或玻片等固相载体上，即制成了组织微阵列(图 17－34)。每个模具可以连续切出许多个组织微阵列。制备好的组织微阵列可与标记探针(同位素、生物素或荧光染料标记)杂交(ISH 及 FISH)或与抗体进行反应(IHC)，在 DNA、RNA 或蛋白质水平检测靶分子，结果扫描后输入计算机用相应的软件进行分析。

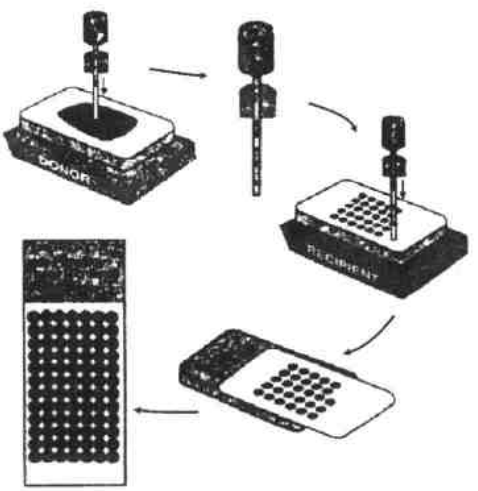


图 17－34　组织微阵列构建示意图

Kononen 应用组织微阵列的方法将 645 种肿瘤组织固定于载体上制成组织芯片，同时检测 *P* 53、雌激素受体（*ER*）、*Myc*、*ERBB*2、*CCND*1、*MYBL*2 等 6 种基因的表达情况。同时根据 *P* 53 和 ER 表达情况，进一步将其分为 4 组检测其他基因的表达情况，对乳腺癌进行更详细的分类及研究。发现 *ERBB*2 和 *myc* 在缺乏雌激素受体的肿瘤中较存在雌激素受体的肿瘤有明显增高，而 *CCND*1 则无明显改变。同时 *myc* 和 *ERBB*2 在 *P* 53 表达阳性肿瘤中有明显增高。

与传统的组织病理学研究方法相比，组织微阵列具有明显的优势：首先，该技术具有组织“扩增”功能。一个标准的组织块约 3～5 mm 厚，理论上可以切上千张 5 μm 厚的切片，但由于技术人员的水平等诸多因素的影响，往往只能切 50～100 张左右。而在构建组织微阵列时，每个组织块可穿刺成 200～300 个柱状物，每个柱状物构成的微阵列又可切成 2 000 张 2.5 μm 厚的薄片，相当于将有限的组织资源“扩增”了 4 000～6 000 倍。其次，组织微阵列中的每个样本操作过程及实验条件完全相同，结果更具可比性，因此有利于实现质量控制及显色方法和结果解释的标准化。第三，低耗费。每个微阵列仅需几微升试剂即可检测几百甚至上千份样本，这样可以节约试剂，降低成本，从而使某些疾病（如肿瘤）的大规模筛查成为可能。第四，有利于对存档的病理检材进行大规模的回顾性研究，这在传统的病理学研究中无法做到。

作为一种全新的组织病理学研究技术，TMA 尚存在着一些问题需要进一步解决。首先是构建微阵列中每个组织片的代表性问题。构建微阵列时，每一样品取样多了，阵列中所排列的总样品数就少，会影响 TMA 的通量；取样少可增加通量，但这么小的组织片能否真正代表原始待检组织值得怀疑。其次是癌前病变研究中的代表性问题。对癌前病变或原位癌，通常只有少量的细胞发生病变，其大小或厚度不足以进行此类研究。炎症或感染性疾病甚至癌组织因具有高度不均一性，同样存在此类问题。另一个潜在的限制是构建微阵列时需使用福尔马林固定的组织，而福尔马林固定的组织不能用于对 RNA 进行研究。目前还没有用冰冻组织构建阵列进行成功研究的报道。对此，用非交联固定剂（甲醇或乙醇）替代福尔马林有望解决。

TMA 可在原位检测大量临床组织样本的分子改变，为翻译基因组研究提供了一个有效的技术平台。由于 TMA 一次可对数千组织样本在 DNA、RNA 或蛋白质水平进行分析，能快速提供关于新候选基因或蛋白的细胞定位、流行病学及临床意义等方面的综合信息使以基因组为基础的靶分子研究更方便、快捷。组织微阵列技术具有显著性加速搜寻分子改变和癌症的临床病理特征间的联系的分子研究的潜能，故对分子和形态病理学家而言，组织微阵列已成为最有前途的研究工具之一，其在癌症研究及其他病理学领域将有极其广泛的应用。目前已有许多用 TMA 进行癌症研究的报道（表 17－6）。

表 17－6　TMAs 在癌症研究中的应用

组织类型	*n*	方法	分子靶标	目的	时间
乳腺癌	645	FISH、IHC、mRNA ISH	*MYC*、*HER*2、*CCND*1、17*q*23、20*q*13、*MYBL*2、*ER*、*P*53	频率、方法比较、分子谱	1998
乳腺癌	557	IHC	*ER*、*PR*、*HER*2	频率、整体比较	1999
乳腺癌	612	FISH、IHC	*HER*2	频率、预后、自动 IHC 记分	2000
乳腺癌	668	FISH	*S*6*K*、*HER*2	频率、预后、cDNA 芯片可靠性	2000
乳腺癌	372	FISH	*RAD*51*C*、*S*6*K*、*PAT*1、*TBX*2	频率	2000
乳腺癌	328	IHC	*ER*、*PR*	方法学、整体比较	2000

续　表

组织类型	*n*	方法	分子靶标	目的	时间
乳腺癌	380	IHC	*ER*、*PR*、*HER2*	方法学比较	2000
乳腺癌	113	IHC	*CCND1*、*MEK*－1 频率、*cDNA*	芯片可靠性	2001
前列腺癌	371	FISH	*NMYC*、*MYC*、*HER2*、*CCND1*	频率、肿瘤预后	1999
前列腺癌	264	IHC	*IGFBP2*、*HSP27*	肿瘤预后、cDNA 芯片可靠性	1999
前列腺癌	458	IHC	*CGA*、*SYN*	频率、方法学比较	2000
前列腺癌	892	IHC	*Ki*－67	频率、种属比较	2000
前列腺癌	632	IHC	*NKX3.1*	频率、肿瘤预后	2000
肾癌	532	IHC	*Vimentin*	频率、cDNA 芯片可靠性	1999
膀胱癌	2317	FISH、IHC	*Cyclin E*	频率、临床相关分析	2000
神经胶质瘤	418	IHC	*IGFBP2*	频率、肿瘤预后、方法学	2000
神经胶质瘤	88	IHC	*Topo* Ⅱ *alpha*	频率、相关分析	2000
神经胶质瘤	160	IHC	*P53*	频率	2000
17 种肿瘤组织	397	FISH	*CCND1*、*CMYC*、*HER2*	不同肿瘤类型的频率	1999
135 种肿瘤组织	4700	FISH	*MB*－*17A*	不同肿瘤类型的频率	2000

蛋白质是细胞中主要的生物效应分子。用 cDNA 微阵列确定细胞内 mRNA 水平不能提供从基因到蛋白质的翻译效率、翻译后蛋白修饰的程度或大量蛋白间相互作用的有关信息。了解不同类型细胞中所有蛋白质的水平、功能活性及其相互作用是“后基因组”时代研究的主要内容。这对疾病的发生、发展极为重要。

经典的蛋白质研究技术，如二维电泳仅适合于检测大多数高丰度的蛋白质，而对绝大多数转录调节蛋白和其他可能具有重要作用的稀有蛋白无法检测；抗体免疫技术通量有限，即使使用微量滴定板，每天也只能进行数百上千次分析，且该技术一般只能分析已知的蛋白产物，而蛋白组学研究需同时对大量已知和/或未知蛋白质进行研究，显然传统蛋白研究技术已无法适应这一新的要求。如今，在 DNA 芯片基础上发展起来的抗体微阵列技术，即蛋白芯片，将填补基因组学和蛋白组学之间存在的信息空白。该技术对细胞群、组织或生物体液中众多蛋白的定量研究极其适用，它将数百或数千抗体有序排列在玻片上，与来源于待检和对照样本的荧光标记的蛋白发生反应。微阵列上的抗体与溶液中蛋白的抗原表位结合，根据两样本间特异蛋白的荧光比率进行定量。

大量蛋白质组学筛选的另一途径是使用 Aptamer 阵列。Aptamer 是一种 DNA 或 RNA 的寡聚物，在体外通过指数式富集法配体进化（systematic evolution of ligands by exponential enrichment，SELEX）过程从随机序列文库中产生。Aptamer 一词来源于拉丁词“Aptus”，意思是“适合”。顾名思义，Aptamer 可用于诊断和治疗过程中的分子识别。Aptamer 与抗体不同，但有类似抗体的特性，能高特异性、高亲和力识别其特异的配体，并与之相互作用。理论上，每个分子都有它自己特异的 Aptamer。用 Aptamer 来探测非核酸靶标有以下优点：① 与靶标结合的亲和力和特异性不亚于单克隆抗体，且它所能识别的靶标范围更广，包括毒素及无免疫原性的靶标；② 通过精确的体外选择过程可筛选出具有最高亲和力和特异性的 Aptamer；③ Aptamer 可人工合成且

稳定性好，不易变性和降解，能长期储存；④ 容易标记和操作。蛋白组学研究需分析大量的蛋白，包括已知和未知蛋白，抗体技术显然已无法适应，Aptamer 正好可满足这一需求。与成熟的抗体技术相比，Aptamer 研究仍处于起步阶段，但其发展非常迅速，有关这方面的研究正在广泛展开。

最后，也可将数百或数千个个体蛋白排列在玻片上，用这种阵列去鉴定蛋白质的功能活性、相互作用的蛋白成分或蛋白的小分子抑制剂及其他许多可能的相关特性。与 DNA 微阵列相比，蛋白芯片技术仍处于早期发展阶段，但在蛋白组学这个重要研究领域中其发展将会非常迅速，可以预计在不久的将来，生物学家将会用和如今分析基因组相同的方式对蛋白质组进行分析。

犹如天文学家拥有一台新的超倍望远镜一样，生物学家借助生物芯片技术获得的分子信息可以更好的观察和揭示人体奥秘，有助于在人类健康的认识方面取得突破。同时，该技术可使病理学家们的工作更简单、有趣，而且也可提供大量有价值的信息帮助我们填补对许多疾病认识的空白。

（二）生物芯片技术在法医病理学中的应用

法医病理学是法医学的主干，它是运用病理学的理论和技术研究和解决涉及法律问题的非自然死亡者的死亡类型、原因、征象、机制及方式等问题的一门学科。法医学是一门应用学科，其发展离不开相应基础学科的发展。生物芯片技术推动病理学飞速发展的同时必将极大推动法医病理学的发展。损伤、死亡时间推断及各种猝死是法医病理学研究的重要内容，生物芯片技术在这些方面均具有重大的研究应用价值。

1. 在脑损伤中的应用。颅脑损伤为最常见的一类机械性损伤，是国内外法医病理学及神经外科学中最常见和最主要的死因之一，一直是相应学科基础和临床研究的热点。由于脑损伤形态学变化同机能障碍的不相称性，有时形态改变不显著的脑损伤导致死亡，此时按传统观点诊断死因就会遇到困难。此外，脑损伤时间的推断历来是法医学研究的热点和难点，传统的方法是根据肉眼及组织学的变化来推断损伤时间，由于影响因素较多，其准确性较差。

随着现代免疫学和分子生物学技术的飞速发展，不少人应用免疫组织化学、分子生物学技术（如原位杂交、RT－PCR、Northern、Western）结合图像分析，对脑损伤进行了研究，以期在脑损伤的诊断和损伤的时间鉴定上有所突破。但到目前为止，多为动物实验结果，少部分人体脑损伤材料的研究，由于研究方法、观察指标不同和缺乏系统性、操作繁琐等，距司法鉴定实际应用还有很大的差距。同时，致命性的脑损伤有时常规病理切片上并无明显变化，单靠常规病理观察难以鉴定死因。近年来，随着学科的发展和研究的深入，人们对创伤修复过程的分子机制有了一定的了解。CNS 损伤可引起一系列复杂的细胞反应，如损伤后立即释放兴奋性氨基酸，导致神经元死亡；单核细胞从血液中迁移进入毁坏的神经组织，转化成巨噬细胞；胶质细胞和毛细血管内皮细胞增生，胶质疤痕以及神经新芽形成等。这些改变有些不利于损伤组织的修复及神经功能的恢复，如疤痕形成会成为再生神经元建立新突触联系的障碍；有些则对损伤组织的修复及神经功能的恢复有利。CNS 是一个复杂的神经网络，神经元和非神经元细胞相互支持，以维持彼此的存活。CNS 损伤既影响损伤部位的细胞，也影响远离损伤部位的细胞。损伤刺激通过信号传导途径激活细胞内的一些转录因子，如 C/EBP、NF－(B、CREB、STAT3、Krox 等，这些转录因子在核内与特定部位的 DNA 区域结合调节大量因子的转录和表达，包括 NGF、BDNF、NT－3、IL－1、IL－6、HSP70、bFCF、TGF－1、TNF－1，等等。创伤后的细胞反应使得介导和调节创伤修复过程的上述蛋白和因子表达增加，并呈现出规律性的时序性改变，可望成为颅

脑损伤诊断和推断脑损伤时间的新依据。

目前，脑损伤的分子病理研究多采用免疫组织化学、原位杂交、RT－PCR、Western、Northern等技术在蛋白、DNA或RNA水平检测上述大量细胞因子、营养因子、转录因子等的表达，对脑损伤后的损伤反应和修复机制进行研究。现今对脑损伤的一些分子机制的认识均采用这些传统的研究方法获得，取得了不少成绩。这些方法虽然准确、可靠，但检测通量有限，无法同时检测大量基因的表达。而CNS是一个非常复杂的神经网络，脑损伤可引起成千上万的基因同时表达，大量检测这些基因的表达可掌握其中包含的时空信息，有助于了解基因的调控机制。生物芯片技术的出现，使脑损伤的全面、综合分析成为可能。随着人类基因组计划的完成，通过基因芯片技术能确定细胞内所有基因的表达谱。将基因芯片技术与差异表达技术有机地结合，通过对脑损伤后损伤周围各种神经细胞和相应正常对照脑组织基因表达进行差异比较，可对脑损伤的发生、发展的分子机制和调控机理进行详细的研究。

用数千个靶基因（如原癌基因和抑癌基因、细胞凋亡相关蛋白及应急反应蛋白、细胞受体、生长相关基因、转录因子等）cDNA克隆制成cDNA芯片，从人脑损伤组织及对照样本抽提mRNA，用RT－PCR方法标记并纯化。在一链合成中掺入荧光标记dUTP，用Cy3－dUTP标记正常对照脑组织mRNA，Cy5－dUTP标记损伤脑组织mRNA，制成杂交探针。探针与芯片杂交后，用扫描仪提取基因表达的荧光信号强度值，经ImaGene 3.0软件分析Cy3和Cy5两种荧光信号的强度和比值，筛选出与脑损伤显著相关的表达基因（图17－35）。再用这些筛选出的基因制成cDNA诊断芯片，即可用于脑损伤的诊断及损伤时间的推断。

2. 在冠心病猝死诊断中的应用。猝死是貌似健康的人因潜在的疾病或功能障碍所引起的突然的意外的死亡。由于其死亡具有很大的意外性，往往引起怀疑，因此猝死的死因诊断是法医病理学的一个重要内容。在成人猝死中心血管系统疾病占首位，尤其是冠心病。在冠心病猝死者中绝大多数具有较明显的血管、心脏病理改变，诊断无多大困难。但少数冠心病猝死者，未见明显的病理改变，如冠状动脉粥样硬化不明显、心肌也未见明显的梗死灶等。心肌梗死灶一般需经8～12 h方可辨认，若症状发作后8 h内死亡，病理学HE染色往往不能发现心肌梗死的病变，这样给部分冠心病猝死的诊断带来困难。因此，发展更灵敏、准确的诊断方法尤为重要。

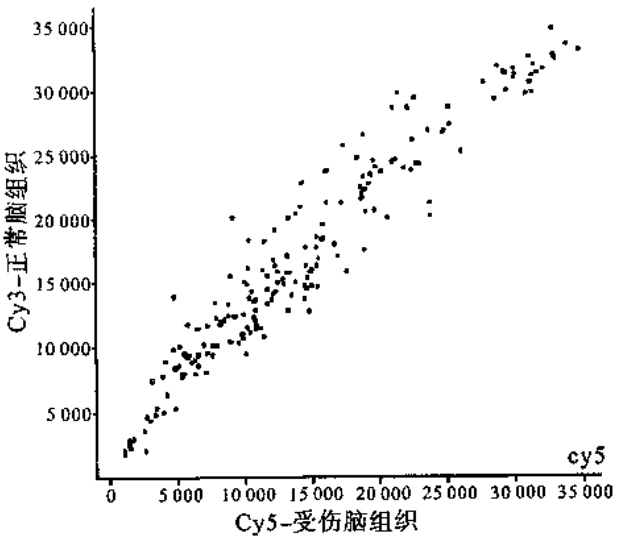


图17－35　不同脑组织样本（对照组/实验组）基因芯片杂交信号强度散点图

冠心病为一多因性疾病，传统的冠心病危险因素如高血压、高血脂、糖尿病、吸烟等只能解释约一半有临床表现的动脉粥样硬化病人，炎症、免疫机制在冠状动脉粥样硬化发生、发展中的作用越来越受到人们的重视。随着研究的深入，人们发现一些主要来源于参与动脉粥样硬化形成的内皮细胞、单核/巨噬细胞、淋巴细胞、平滑肌细胞和血小板等的细胞因子（如心肌营养素－1、PDGF、VEGF、IL－1、TNF、TGF、MCP－1、NOS 等）、粘附分子（如 ICAM－1、VCAM－1、PECAM－1 等）和转录因子（如 NF－B、CREB 等）的表达发生改变，研究它们的表达改变规律有助于了解冠心病发生、发展的分子机制，同时可望为冠心病的分子病理诊断提供客观依据。

Zhang 等（2001 年）用含 4 500 个基因的 cDNA 微阵列分析了尼古丁对人类冠状动脉内皮细胞中基因表达的影响，发现大量基因的表达发生改变，如磷脂酰肌醇磷酸激酶表达增加，二酰甘油激酶表达下降，转录因子 CREB 和 NF－B 表达增加等。Coombes 等（2001 年）用含 268 个基因的 cDNA 阵列分析人类内皮细胞对肺炎衣原体感染的反应，观察到感染后 2 h *IL*－1、*IL*－8、*MCP*－1、*BFGF*、*HBEGF*、*PDGF*－B 等基因表达增加，同时细胞内激酶和细胞表面受体 mRNA 水平也升高。这些资料有助于了解抽烟及衣原体感染在冠状动脉粥样硬化和冠心病发病机制中的病理生理作用。

生物芯片技术的发展使全面分析研究冠心病相关基因的表达成为可能，预计不久的将来在全面了解冠心病发病机理的基础上，会研究出简便、准确的冠心病诊断方法，从而解决法医学上冠心病猝死诊断这一难题。

3. 其他猝死的诊断。婴幼儿猝死综合征（SIDS）和青壮年猝死综合征（SMDS）的诊断一直是法医病理学的难点。这类猝死者系统尸体剖验和有关辅助检查，查不出足以说明死因的器质性病变，也无中毒或外伤等暴力致死证据。这类猝死可能不是单一的病因引起的，而是包含多种病因。没有明确客观的形态学诊断指标，给其诊断带来了困难。现代生物学，尤其是生物芯片技术的发展给这类猝死的诊断带来了曙光。从基因水平研究这类猝死发生的机理及其改变，可望给其诊断找出一些客观依据。

三、生物芯片在痕量物证及毒品和毒物分析中的应用

（一）痕量物证的分析及毒品、毒物的检验技术

痕迹检验中对爆炸残留物、枪击残留物、指甲刮痕中有机物残留及染发成分等的检测与药品及毒品检验的方法相同，这里一并叙述。分析的常规仪器有光谱法（紫外分光光度法、红外光谱法、荧光光谱法、原子吸收光谱法）、色谱法（薄层色谱法、气相色谱法 GC、高效液相色谱法 HPLC）、质谱法（MS）、气相色谱－质谱（GC－MS）联用技术等。近年来，高效毛细管电泳（CE）已成为发展最快的分析技术之一。CE 是经典电泳技术和现代微柱液相分离技术相结合的产物。CE 和 HPLC 均属于高效的液相分离技术，仪器均可自动化，且均具有不同的分离模式。但 CE 分析时间常不超过 30 min，比 HPLC 快；CE 的柱效比 HPLC 要高得多；CE 所需样品量为 nl 级；流动相用量也只需几毫升。CE 有多种分离模式，迄今以毛细管区带电泳（CZE）及胶束电动毛细管色谱（MECC）在毒品、毒物分析中的应用最多。现已有大量报道 CE 用于阿片类、巴比妥类毒品药物的研究报道。随着检测技术的进一步发展，检测灵敏度的进一步提高，CE－MS 联用技术的发展成熟，CE 技术在法庭毒物分析领域将显示更大的优越性。

（二）生物芯片在痕量物证分析及毒品、毒物中的应用

痕量物证分析包括石化产品、核爆炸物分析、指纹分析、弹药残骸分析等，毒品一般指大

麻、吗啡、可卡因、安非他明等。传统方法主要是 HPLC、UV、GC－MS、X－ray 等，所用仪器价格昂贵，操作过程复杂，速度较慢。毛细管电泳芯片的出现使痕量物证分析变得非常容易和快速。研究表明，运用芯片毛细管电泳技术可以在数秒或数分钟内完成对待检样品的分离，并可以高通量地同时对多个样品进行分析，进行快速准确的现场分析，在最短的时间内为侦察人员提供第一手资料，对于刑事案件的快速侦破具有重要意义。

四、前景与展望

用分子微阵列技术分析疾病样本获得的信息必须在组织和细胞形态学的层面加以解释。正常组织和疾病组织细胞构成的复杂性，使得分子微阵列资料的解释复杂化。而且，许多微阵列实验研究的是模型系统，如细胞系、异种移植或动物肿瘤模型，还必须通过临床研究加以印证。因此，该技术从基础研究到走进诊断病理学实验室真正进入实际实用尚需时日。

微阵列技术仍然处于发展阶段，它也面临着一些问题尚待解决，如高耗费、需要特殊的专业技能、需要不同学科间的合作、大量的资料的处理问题及所得结果的生物学意义和临床应用的不确定性等等。尽管如此，生物芯片技术的发展方兴未艾，各种用于病理检测的工程性生物芯片，如盘式免疫分析微阵列(disc－based immunoassay microarrays)以及用于脑损伤诊断及时间推断的微阵列正在构建研究之中。微阵列技术的应用对诊断病理学实践带来的冲击可能远比当初 IHC 给病理学带来的改变更大得多。未来的后基因组学研究将把人类医学带入崭新的时代：将由主要依赖经验、形态学改变转向以特异的分子病理学为依据。不仅有一天微阵列分析的结果在治疗的指导和预后的评估方面会压倒组织病理学诊断，而且可能在不久的将来，组织病理学的本质在基因组和基因表达谱分析的基础上将会被重新定义。

生物芯片技术在后基因组研究中将起到举足轻重的作用。高密度的基因芯片已经在功能基因组研究中发挥了独特作用，实现了基因表达谱分析的微量化、集成化和连续化。生物芯片检测自动化程度的不断提高，将出现从样品(如血液、组织等)预处理到数据分析全过程在一张芯片上完成的高技术集成系统－芯片实验室(lab－on－a－chip)。1998 年 6 月，美国 Nanogen 公司的程京博士和他的同事们首次报道了用芯片实验室所实现的从样品制备到反应结果显示的全部分析过程。他们用这个装置从混有大肠杆菌的血液中成功地分离出了细菌，在高压脉冲破胞之后用蛋白酶 K 孵化脱蛋白，制得纯化的 DNA，最后用另一块电子增强的 DNA 杂交芯片分析证实提取物是大肠杆菌的 DNA。这是向缩微实验室迈进的一个成功的突破。目前，含有加热器、微泵微阀、微流量控制器、电子化学和电子发光探测器的芯片已经研制出来了，而且，也出现了将样品制备、化学反应和分析检测部分结合的芯片。

缩微芯片实验室的核心技术是在芯片上含有 DNA 样品制备、纯化与检测等微电子结构。由于其具有精密的微电子制备和检测系统，不仅所需样品量大为减少，而且灵敏度提高，因而可以快速、精确地完成从样品制备到反应结果显示的全部分析过程。可以有效克服人工操作的实验误差，系统稳定性能极佳。相信包含全部实验过程的缩微芯片实验室将很快出现，人们可以在一个封闭的系统内快速完成从原始样品到测试结果的全部实验步骤，进一步结合卫星传输和网络生物信息学的技术资源，真正实现高通量、一体化、移动性的“未来型掌上实验室”的构想。从而建立全世界联网的司法鉴定实验室，为司法科学的发展做出贡献。

(杨庆恩　黄代新　陈　慧)

第十八章
蛋白质芯片

内容提要

蛋白质组的研究要同时处理成千上万种蛋白质，而蛋白质芯片是用于蛋白质功能研究及相互作用分析的生物芯片，这种高通量技术提供了在同一时间分析各个蛋白质组的可能，是蛋白质组学时代最有启发意义的技术工具。本章节主要介绍蛋白质芯片产生的背景、基本原理、在医药学上的应用现状与进展及其发展趋势与前景。

后基因组时代着重于从整体的角度系统分析生物体的功能机制。蛋白质芯片是一种高通量的蛋白功能分析技术。它可以用于构建蛋白质表达谱，进行抗原—抗体筛选，药物靶点筛选，蛋白质－蛋白质交互作用筛选等，是对蛋白质组进行功能分析的新技术，可广泛用于医学、生物学、食品安全检测、法医学、环境监测等领域。

第一节　蛋白质芯片产生的背景

一、后基因组时代

“功能基因组学”在20世纪80年代还被认为是超现实主义者、“科学前卫派”们口头的词汇，进入21世纪后却已成为生命科学中最具活力和发展潜力的领域。伴随着人类基因组计划的顺利实施，人类基因组序列草图的完成代表着一个全新的纪元——“后基因组时代”的到来。HGP的实施使人们获得大量DNA序列片段，促使人们去探究其中的奥秘，其中所蕴藏生物信息亦即功能基因组数据。科学家不仅希望阐明每个基因的具体功能，更希望通过新的方法完全澄清各种组织与器官在基因水平上的调控机制，最终能接近生命体的本质。在功能基因组时代，我们必须也正在经历一个转变，即从单个基因功能的研究转变为从整体的角度高通量地、系统地进行基因组功能研究。生物学家总是希望在同一时间平行地分析某一生物系统所有功能基因的表达谱。与此相适应，出现了高通量技术（high－throughput technique），它们的共同特点是可以同时研究大量基因及其表达产物，具有信息量大、自动化、多层次、高效率的优点，主要包括微阵列（microarray）或生物芯片（biochip）技术。后基因组时代的相关研究不仅对了解生命的本质具有重要意义，而且会在医学，农学，动物学，植物学，食品卫生学，药学等等领域产生巨大而广泛的应用前景。这将是科学技术的又一次划时代的革命。

二、蛋白质组学的兴起与发展

20世纪末期出现的DNA芯片为描绘不同组织的mRNA、cDNA表达谱提供了可能，已经得到广泛的应用。但是我们知道，蛋白质是生命活动的执行者和体现者，是生物体最重要的结构成分、催化剂和信号传递分子。蛋白质是一切生命活动的基础，蛋白质的数量和质量是受基因表达的调控，但是蛋白质本身亦有其特殊的活动规律，随着生命活动的进程表现出动态的紧密协调变化；蛋白质在合成后具有相对独立的修饰、运转和相互间作用能力；同时还具有对外界因素发生反应的能力。况且基因的主要功能也是通过其表达产物——蛋白质来实现的。众所周知，基因在转录时，一个基因可以多种mRNA形式剪接，并且同一蛋白质可能以多种形式进行翻译后的修饰，这就造成一个蛋白质组不是一个基因组的直接产物，蛋白质组中蛋白质的数目可以超过基因组的数目。蛋白质表达谱比mRNA表达谱更能真实地反映生物体的功能机制。传统的对单个蛋白质进行研究的方式已无法满足后基因组时代的要求。这是因为：① 生命现象的发生往往是受多因素影响的，必然涉及多个蛋白质；② 多个蛋白质之间的相互作用是交织成网络的，或平行发生，或呈不同级结构间的交互因果关系；③ 在执行生理功能时蛋白质的表现是多样的、动态的，并不像基因组那样基本不变。要对生命的复杂活动有全面和深入的认识，必然要在整体、动态、网络的水平上对蛋白质进行研究。

因此，只有从整体的角度对所有蛋白质的总和进行研究，才能更有效地掌握生命活动的规律，进而接近生命的本质。在这样的情况下，“后基因组时代”的“宠儿”——蛋白质组学应运而生。

蛋白质组(Proteome)是于1994年首次创造的一个术语，而蛋白质组学(Proteomics)是指对一个给定的细胞、组织、有机体在一定时间内表达的所有蛋白质(总蛋白质)，即蛋白质组进行系统的、全面的研究。蛋白质组学的第一篇原始论著发表于1995年国际上并不出名的《电泳》杂志。短短几年的发展，蛋白质组学已经从建立数据库走向解决生命科学重大问题，研究生命活动机制的强有力手段。通过蛋白质组学的研究，科学家们能够发现每种疾病在不同的发病阶段蛋白质水平上发生的变化，对正常个体及病理个体间的完整、动态的蛋白质谱的分析比较，可以找到某些“疾病特异性蛋白质分子”，进而发展成为临床早期诊断指标，形成未来分子诊断学、分子治疗学的理论和应用基础，它们还可能成为新药物设计的分子靶点。当然，蛋白质组学发展和应用还有很多方面，但它对国际生物制药业的推动无疑是巨大的。蛋白质组市场在2000年产生了9.63亿美元的收入，到2006年预计可达56亿美元，其在理论研究和应用研究方面都呈几何级数增长的发展趋势。

三、传统技术的改造和新技术的产生

蛋白质组学的发展既是技术所推动的，也受到技术的限制。蛋白质组学研究成功与否，很大程度上取决于其技术方法水平的高低。蛋白质研究技术远比单纯基因技术复杂和困难。不仅是因为氨基酸残基种类远多于核苷酸残基(20/4)，而且蛋白质有着复杂的翻译后修饰，如磷酸化和糖基化等才能成为有功能活性的蛋白质，而且蛋白质的变化多，使分离和分析蛋白质的困难增大。此外，通过表达载体进行蛋白质的体外扩增和纯化制备大量的蛋白质并非易事。蛋白质组学的兴起对技术有了新的需求和挑战。蛋白质组的研究实质上是在分子水平上对蛋白质进行大规模的平行分离和分析，往往要同时处理成千上万种蛋白质。因此，发展高通量、高灵敏度、高准确性的研究技术平台是现在乃至今后一段时间内蛋白质组学研究的主要任务。

当前国际蛋白质组研究技术平台的发展趋势表现在以下几个方面：

(一)样品的制备和分级技术

蛋白质组研究中的样品制备通常可采用细胞或组织中的全蛋白质组分进行蛋白质组分析。为了达到这个目的也可以进行样品预分级，即采用各种方法将细胞或组织中的全体蛋白质分成几部分，分别进行蛋白质组研究。样品预分级可使低表达的蛋白质得到富集，可以提高低丰度蛋白质的上样量和检测，主要预分离的方法包括电泳或色谱预先分离、蛋白质亲和富集以及亚细胞分离、细胞器定位分级等。①液相等电聚焦，可以将复杂的蛋白质混合物在一定的pH值范围分离，将其浓缩在相对较窄的范围内；②吸附色谱，常用的有离子交换色谱(IEC)，疏水相互作用色谱(HPLC)，反向色谱和亲和色谱，体积排阻色谱(SEC)等，色谱与2DE结合可用于对大量蛋白质的快速纯化；③亚细胞分离，可对特定的蛋白质组进行分离纯化，减少了蛋白质的分散性和复杂性，主要步骤包括破坏细胞组织和分离不同器官。

最近在组织水平上的蛋白质组样品制备方面也有新的进展，如采用激光捕获显微解剖技术(laser capture microdissection，LCM)方法分离癌变上皮类细胞。基本方法是：在显微载玻片上用一种塑料膜覆盖住组织切片，用激光束活化并吸附到解剖区域，组织切片部分热溶于胶片上，能从载玻片上剥离下残余组织。

(二)样品分离和分析技术

利用蛋白质的等电点和分子量通过双向凝胶电泳的方法将各种蛋白质区分开来是一种很有效的手段。它在蛋白质组分离技术中起到了关键作用。其基本的技术包括在第一向等点聚焦中使用IPG干胶条，将蛋白质成分按其等电点经分离后均匀的分布于不同的pH梯度中(等电点处)然后再进行SDS－PAGE第二向电泳分离。经过双向电泳后，样品中的蛋白质分辨率得以极大提高。每一个细胞或组织都会从中得到详尽的"蛋白质指纹图"，而进行特征性蛋白质组研究。

如何提高双向凝胶电泳的分离容量、灵敏度和分辨率以及对蛋白质差异表达的准确检测，是目前双向凝胶电泳技术发展的关键问题。做过双向凝胶电泳的人一定会抱怨它的繁琐、不稳定和低灵敏度等缺点。发展可替代或补充双向凝胶电泳的新方法已成为蛋白质组研究技术最主要的目标。目前，二维色谱(2D-LC)、二维毛细管电泳(2D-CE)、液相色谱—毛细管电泳(LC-CE)等新型分离技术都有补充和取代双向凝胶电泳之势。其他的方法还有第一维电泳采用窄pH梯度胶分离以及开发与双向凝胶电泳相结合的高灵敏度蛋白质染色技术，如新型的荧光染色技术。

(三)灵敏的检测方法

蛋白质混合物经分离后采用灵敏的、合适的检测方法是双向电泳后分离的关键步骤。常用的检测标记物有三类：放射性核素、有机染料和荧光素。

质谱(MS)技术是目前蛋白质组研究中发展最快，也最具活力和潜力的技术。它通过测定蛋白质的质量来判别蛋白质的种类。当前蛋白质组研究的核心技术就是双向凝胶电泳— 质谱技术，即通过双向凝胶电泳将蛋白质分离，然后利用质谱对蛋白质逐一进行鉴定。对于蛋白质鉴定而言，高通量、高灵敏度和高精度是三个关键指标。一般的质谱技术，难以将三者合一，而最近发展的质谱技术可以同时达到以上三个要求，从而实现对蛋白质准确和大规模的鉴定。以质谱技术为核心，开发质谱鸟枪法(shot gun)、毛细管电泳 —质谱联用(CE—MS)等新策略，直接鉴定全蛋白质组混合酶解产物。随着对大规模蛋白质相互作用研究的重视，发展高通量

和高精度的蛋白质相互作用检测技术也被科学家所关注。实际上质谱相关技术由来已久，但是没有应用到生物大分子研究的主要原因在于以前的质谱技术电离方式产生大量碎片离子，结果图谱多得无法解释。20 世纪 80 年代后期出现了液相电喷雾质谱和基质辅助的表面增强激光解吸附电离化飞行时间质谱(surface enhanced laser desorption ionization - time of flight，SELDI - TOF - MS)，具有高敏感性、特异性和快速的特点，这一质谱技术的突破以及各种分析软件出现与更新，使 2 - DE 分离的成千上百个蛋白点得到鉴定变为现实。利用从原始样品亲和吸附特定蛋白进行分离，以进行鉴定，可高速的进行活性蛋白质组的分析。所以 2 - DE 技术、质谱技术的联姻为蛋白质组从结构上研究蛋白质提供了强大的技术支持。也成为蛋白质芯片技术的基础。

(四)蛋白质组生物信息学及蛋白质组数据库检索

随着越来越多的蛋白质混合物被分离、鉴定，产生的相应数据越来越多，为了确定已知蛋白是否为新蛋白，通过数据库的检索比较是非常重要和必需的，是蛋白质组研究水平的标志和基础。瑞士的 SWISS PORT 拥有目前世界上最大、种类最多的蛋白质组数据库。丹麦、英国、美国等也都建立了各具特色的蛋白质组数据库。生物信息学的发展已给蛋白质组研究提供了更方便、有效的计算机分析软件。特别值得注意的是，蛋白质质谱鉴定软件和算法发展迅速，如 SWISS PORT、Rokefeller 大学、UCSF 等都有自主的搜索软件和数据管理系统。最近发展的质谱数据直接搜寻基因组数据库，使得质谱数据可直接进行基因注释，判断复杂的拼接方式。随着基因组学的迅速推进，会给蛋白质组生物信息学研究提供更多更全的数据库，也为蛋白质芯片的分析提供了有力的比照数据系统。

在基础研究方面，近两年来蛋白质组研究技术已被应用到各种生命科学领域，如细胞生物学、神经生物学、发育生物学等。在研究对象上，覆盖了原核微生物、真核微生物、植物和动物、人等范围，涉及到各种重要的生物学现象，如信号转导、细胞分化、蛋白质折叠等等。未来蛋白质组学的研究领域将更加广泛。在应用研究方面，蛋白质组学将成为寻找疾病分子标记和药物靶标最有效的方法之一。在对癌症、早老性痴呆等人类重大疾病的临床诊断和治疗方面，蛋白质组技术也有十分诱人的前景。在技术发展方面，蛋白质组学的研究方法将出现多种技术并存，各有优势和局限的特点，而难以像基因组研究那样形成比较一致的方法。除了发展新方法外，更强调各种方法间的整合和互补，以适应不同蛋白质的不同特征。蛋白质芯片这种高通量技术提供了在同一时间分析各个蛋白质组的可能，是蛋白质组学时代最有启发意义的技术工具。例如，通过比较正常状态与疾病状态的蛋白质组，科学家能够更好地理解信号转导、物质及能量代谢通路。许多科学家认为蛋白质芯片将会缩小存在于基因组与蛋白质组之间的信号鸿沟，成为生命科学中最令人激动的前沿。

第二节　蛋白质芯片的基本原理

蛋白质芯片(protein chip)，狭义的表达是蛋白质微阵列(protein microarray)，是用于蛋白质功能研究及相互作用分析的生物芯片，采用原位合成、机械点样或共价结合的等方法将多肽、蛋白、酶、抗原、抗体固定于硅片、玻璃片、塑料片、凝胶、尼龙膜等固相介质上形成的生物分子点阵，见图 18 - 1。在待分析样品中的生物分子与蛋白质芯片的探针分子发生杂交或相互作用或其他分离方式分离后，利用激光共聚焦显微扫描仪对杂交信号进行高通量检测和分析，是将整个蛋白质水平的相关生化分析过程集成于芯片表面，从而实现对多肽、蛋白质及其他生物

成分进行高通量检测。

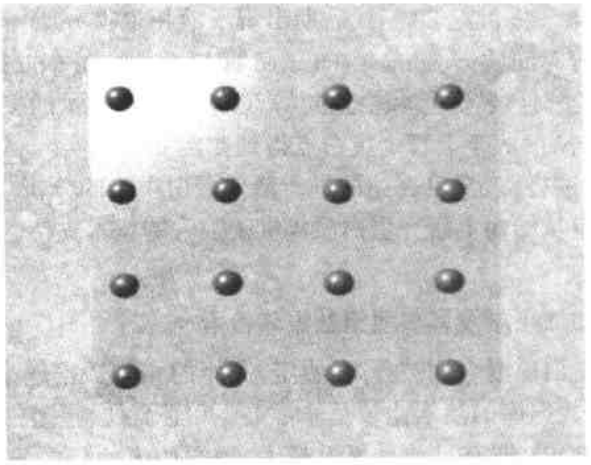


图 18-1 蛋白质芯片

一、蛋白质芯片特点

蛋白质芯片的特点为:① 高通量,高效率;②蛋白质芯片是一个动态的技术流程概念,不局限于某个固相有形物;③ 适用于包括组织、细胞系、体液在内的多种生物样品;④ 芯片上的样品包含针对信号传导、癌症、细胞周期调控、细胞结构、凋亡和神经生物学等广泛的生物功能的相关蛋白,跨度大、适用范围广;⑤ 芯片上的结合物分别经过不同方法检测,灵敏度高达 pg/mL;⑥ 开放性的芯片平台设计,可以用各种型号 DNA 芯片荧光扫描仪进行检测;综合上述特点,决定蛋白质芯片有着广泛的应用潜力和前景。

二、蛋白质芯片的分类

蛋白质芯片按照不同的标准可进行不同的分类:

(一)按样品结合方式分为化学型和生物化学型

化学型蛋白质芯片的构想来源于经典色谱(反相层析、离子交换层析、金属螯合层析等)的介质。通过介质的疏水力、静电力、共价键等结合样品中的蛋白质,然后经特定的洗脱液除去杂质蛋白质,而保留感兴趣的蛋白质。这种芯片特异性较差。

生物化学型蛋白质芯片则是把生物活性分子(如抗体、受体、配体等)通过生化反应结合到芯片表面,用于捕获样品中的靶蛋白。由于生物化学型芯片具有高度的特异性及生物活性分子的多样性,其应用范围和前景明显优于化学型蛋白质芯片。

Ciphergen Biosystems 公司是最早提出上述分类的。

(二)按点样蛋白质有无活性功能分为无活性的(non-living)和活性(living)芯片

无活性的芯片是将已经合成好的蛋白质点在芯片上,有活性的芯片则是指点在芯片上的样品是活的生物体(如细菌),在芯片上原位表达蛋白质。活性芯片可以提供模拟的机体内环境,对于蛋白质功能分析更为有利。

无活性的蛋白质芯片制作方式主要分 3 类:①原位合成。Ceysen HM 等人采用荧光甲氧羰

基氨基酸保护策略，由 C 端向 N 端，共合成了 208 个六肽排列在直径为 4mm 的聚乙烯膜上，以研究一个由 213 个氨基酸组成的蛋白质的活性域结构。②点合成。Krame A 等人利用该法在纤维素膜上合成多肽文库，纤维素膜上有丰富的羟基，对于多肽的合成是十分有利。③光蚀刻术。基因芯片的创始人 Fodor SPA 采用光蚀刻术固相合成的多肽芯片密度为每平方厘米 25 万点，多肽的平均长度为 5 肽。

活性的蛋白质芯片的制作：Uetz P 等人利用酵母细胞构建了第一个全基因组的活性蛋白质芯片。芯片主要用来筛选蛋白质与蛋白质之间的相互作用，包含了 5 800 个酵母克隆。活性蛋白质芯片通过一次实验就能拿到上千种未知蛋白质的功能线索，例如通过将细胞裂解液中蛋白成分做成芯片，科学家很快就可以查明某一开放阅读框架（ORF）是否表达一种分泌蛋白抑或是一种胞内酶。此外，还可以在芯片上进行活细胞的遗传性状筛选。

（三）按作用和点样的顺序分为蛋白质表达芯片（protein expression chip，PEC）和蛋白质功能芯片（protein function chip，PFC）

PEC 能令人通过观察在不同条件下哪种蛋白质被合成而获得基因表达的详细信息；而 PFC 则能分析众多蛋白质的性质。PEC 的目标与 DNA 芯片相类似，它将大量检测用的分子（常为抗体）固着在芯片表面上而进行微阵列式排列，故加样后就能查明样品中有无这些分子的靶标。目前虽尚不能在芯片上装载有能检测每种蛋白质的文库，但即使部分地做到能检测蛋白质亚群（subset），在生物学研究和新药开发中也有着重要的使用价值，特别是当与较传统的蛋白质功能分析和以核酸为基础的基因表达分析技术综合应用时。

PFC 与 PEC 有许多共同特点，即它们都需要有大量不同分子在芯片表面各分散的点上固着，随后加样进行检测；但不同的是 PFC 表面上固着的却是待测的蛋白质分子，随后再用其他蛋白质、药物和生化试剂来检测各待测分子与它们结合的情况并了解其性质。因为待检测蛋白质的最确定和直接的性质表现为它的结合特异性，即分子间的某种相互作用。随着微射流技术（microfluidics）和其他微处理技术的改进，也有可能做到查明在微阵列上固着的蛋白质所介导的生物学反应。

（四）按检测方法进行分类

较有代表性的有单元芯片、中科院力学所的多元蛋白质光学芯片和美国的 SELDI 质谱芯片等。它们的共同特点都是将生物分子作为配基，固定在固体芯片表面或表面微单元上，以单一、或面阵、或序列式。利用生物分子间的特异结合的自然属性，待测分子与配基分子在芯片表面会形成生物分子复合物。然后，检测此复合物的存在与否，达到对蛋白质的探测、识别和纯化的目的。以上不同技术的差异仅在探测方法的不同。BLACORE 技术利用表面等离子体共振技术检测芯片，进行单一蛋白质检测；多元蛋白质光学芯片是光学成像法，可以同时检测多种混合的蛋白质；SELDI 技术则采用质谱法，以时间顺序检测序列蛋白质。

（五）按固相载体片基种类进行分类

平玻板型（plain－glass slide）、三维胶体填充板型（3D gel pad chip）和微孔板型（microwell chip），见图 18－2 和表 18－1。

（六）按工作原理分为探针型和凝胶电泳型芯片

探针型类似于 DNA 芯片，即在固相支持物表面高密度排列的探针蛋白质点阵，可特异捕获样品中的靶蛋白，然后通过检测器对靶蛋白进行定性或定量分析。另一种就是微型化的凝胶电泳板。在电场作用下，样品中的蛋白质通过芯片上的孔道分离开来，经喷雾直接进入质谱

仪进行检测，以确定样品中蛋白质的分子量及种类。就应用前景而言，人们更关注第一种。

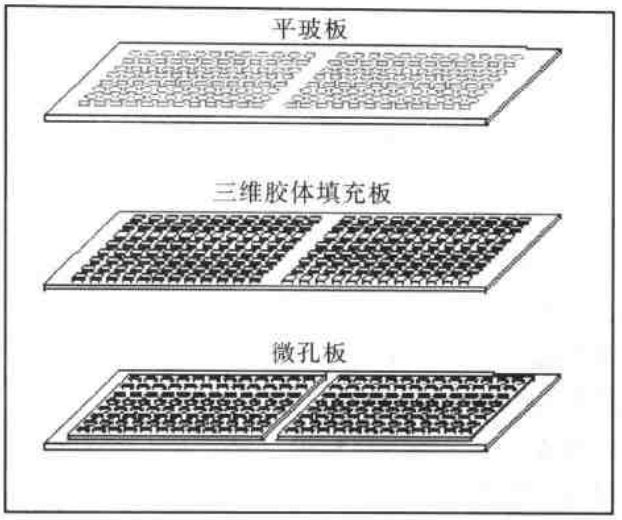


图 18－2　不同板型的蛋白质芯片

表 18－1　不同类型蛋白质芯片的性质比较

玻璃平板	胶体模板	微孔板
可与标准微阵列仪及探测设备兼容	可与标准微阵列仪及探测设备兼容	可与标准微阵列仪及探测设备兼容，但需要联合使用
样品高挥发	减少样品挥发	减少样品挥发
对复合反应效果弱	可进行溶解反应，但对材料需较长洗脱时间	广泛用于溶解及复合反应
成本不贵	成本贵，需光制版技术	成本低
可能会有交叉污染	无交叉污染	无交叉污染

三、蛋白质芯片制作和检测的基本流程

蛋白质芯片的制作，常规的方法先是在一种固相载体上按照预先设计的方式固定大量蛋白质（抗原或抗体），形成蛋白质的阵列，即蛋白质芯片。实验时，往芯片上加入带有特殊标记的蛋白质分子（抗体或抗原），两者结合后，通过对标记物的检测来实现对抗原或抗体确定的互检。

常规的蛋白质芯片制作与检测的流程：

（一）制备抗体文库

如利用噬菌体展示技术。

（二）选择固相载体片基

片基呈薄片状，外观可做成长条状、圆形或椭圆形等不同形状，经特定处理后承载吸附有

关的生物制剂。常用的材质有硅、云母、各种膜片等。各公司制作芯片采用的载体材料基本相差不大。

(三)制备探针蛋白质,点样

对蛋白质芯片来说,这种探针包括特定蛋白质的抗体或受体,结合某些阳离子或阴离子的化学基团、亲水或疏水物质、酶、免疫复合物等,用化学固定技术固定在载体表面上,形成一个个点状芯池(注:一个芯池指一个蛋白质样品的加入孔)。单克隆抗体具有高度的特异性和亲和性,是一种比较好的探针蛋白,用其制作的芯片主要用于检测蛋白质表达丰度及确定新的蛋白质。噬菌体抗体库技术是典型的制备蛋白质探针的技术,其他的蛋白质文库,如全合成人重组抗体库、噬菌体肽库、噬菌体表达文库等,也可用来制备探针蛋白。通过 2 - DE 电泳分离技术分离得到的蛋白质作为抗原,把抗原与固相化的抗体库孵育,即可得相应的单克隆抗体。

由于蛋白质具有易于变性、不易与固体表面结合、在固相支持物表面上不易折叠成有功能的结构而丧失其活性的特点,因此以前的蛋白质多点阵制作过程复杂,仍局限于低通量结构微阵列范畴。在将蛋白质固定在固体支持物表面时,必须保持蛋白质的立体构象,才能使蛋白质探针具有生物活性。

Arenkov 等把探针蛋白固定于微型聚丙烯酰胺凝胶块中构建蛋白质芯片,利用电泳原理捕获样品中能与探针蛋白特异结合的靶蛋白。由于无法保持固体表面蛋白质的活性,使得这些蛋白质芯片无论从制作还是应用方面都要比 DNA 芯片复杂得多。

化学生物学家 MacBeath C 和化学家 Schreiber S 联手解决使蛋白质粘附于固相载玻片上而保持活性这一关键技术突破之后,才真正使印迹蛋白微点阵用于高通量功能研究成为现实。这一技术突破的关键在于对载玻片的处理。用含醛的硅烷试剂处理玻片,并用含有 30% 甘油的磷酸盐缓冲液溶解蛋白质后,再用 Affymetrix 生产的 GMS417 阵列点样机将微量蛋白质点在玻片上,这样,蛋白质的伯氨基与醛形成希夫式碱而将蛋白质以共价结合的方式附着在玻璃表面,再用 BSA 处理载玻片表面,然后用 N,N′二琥珀酰胺基碳酸盐溶液(N,N′ - disuccinimidyl-carbonate)活化 BSA 的赖氨酸、谷氨酸残基,使其易与印迹蛋白的表面氨基作用形成共价脲连接或酰胺连接,待固定的肽或小分子蛋白则可在载玻片 BSA 单层分子表面保持立体构象,易于与溶液中的大分子起反应,这样就可以与蛋白质进行杂交了。研究者或用荧光基因标记蛋白,或用 ATP 标记的激酶底物,都可与膜上的蛋白进行反应。最终他们将一个有 1 万个点的芯片与一个 FRB(FKBP12—rapamycin binding protein)结合蛋白和 G - protein 结合蛋白探针进行杂交。发现在一大片 G - protein 点的中央单独的一个 FRB 的点被区分了出来。

有人用微细加工的聚丙烯酰胺捕获蛋白质,并用微电泳的方法加速蛋白质的扩散,从面将蛋白质固定在芯片片基表面。因聚丙烯酰胺凝胶中含有水分,故可使蛋白质保持活性。

制备探针蛋白时常采用直接点样法,以避免蛋白质的空间结构改变,保持它和样品的特异性结合能力。现已有预制好已知探针的多种芯片系统出售,如微阵列式、微孔板式、凝胶块状等类型。一般来说,一块载体上根据需要固定的芯池可多可少,研究用芯片的芯池数目较少,大多都在 6 ~ 10 个左右,多者可达 10 多个;但供规模生产用的芯片,其芯池数量要多得多。现有一种构象型蛋白质芯片,其芯池数目高达 1 600 个/cm^2,呈微距阵排列,点样时须用机械手进行,可同时检测数千个样品。使用时先将需要检测的含有蛋白质的标本如尿液、血清、精液、组织提取物等按一定程序做好层析、电泳、色谱等前处理,然后在每个芯池内点入需要的种类,一

般样品量只要 2～10 μL 的微量即可，根据测定目的不同，可选用含有不同探针的芯片。让标本在每个芯池中与特定的探针结合或与其中含有的生物制剂相互作用一段时间，然后洗去未结合的或多余的物质，将样品固定一下等待检测即可。

(四)制备待检样品(蛋白质、多肽、酶等)

Von Fggeling E 等人将显微切割技术与蛋白质芯片结合用于肿瘤研究，比较正常组织与肿瘤组织的蛋白质表达谱差异。首先用显微切割术获取纯的肿瘤与正常组织，提取组织中的蛋白，与蛋白质芯片反应，洗脱掉未结合上的蛋白质，利用配套的分析系统检测芯片，作出蛋白质表达谱。显微切割技术能够提供来源同一的组织，非常适合于样本量要求不多的芯片分析与质谱分析，这是二维电泳无法满足的，因为二维电泳所需的样本量比较大。

(五)对样品进行标记

以质谱技术为基础的直接检测法不需标记，标记样品中的蛋白质主要是为随后采用 CCD (charge－coupled device)照相技术及激光扫描系统等对信号检测，标记物主要是荧光物质或同位素。

(六)与芯片进行抗原抗体反应

主要是酶联免疫吸附(enzyme－linked immunosorbent assay，ELISA)，再经洗脱，只有那些与芯片上的抗体发生反应的蛋白质能粘在芯片表面。

Mendoza LG 等人将 ELISA 运用于高通量的蛋白质芯片中。他们使用的是 96 孔的微量滴定板。每个孔中点上 144 个样(36 个样本，重复 4 次)，ELISA 反应照常规进行。在点样时采用了机器手，用毛细吸管来转运蛋白质，精密度非常高，能够将交叉反应降低到最低限度。

(七)蛋白质的分离和检测

蛋白质测定所使用的方法主要是在质谱(MS)分析的基础上发展起来的新技术，常用且比较成熟的有基质辅助的激光解吸附电离化飞行时间质谱(matrix－assisted laser desorption ionization－time of fight－MS，MALDI－TOF－MS)，以及在它基础上改进后实行的表面加强的激光解析电离化飞行时间(SELDI－TOF－MS)。其原理是利用激光脉冲辐射使芯池中的待分析样品解离形成荷电离子，由于具有不同质荷比，这些离子在仪器场中飞行的时间长短不一，由此绘制出一张质谱图，该图经计算机软件处理得到模拟谱图，同时直接显示样品中各种蛋白质的分子量、含量等信息，若将它与正常人或某种疾病患者的谱图，甚至数据库中的谱图进行对照，还能够最终发现和捕获新的特异性相关蛋白及其特征。整个测定过程一般可以几十分钟内就全部完成，且方法敏感，特异性高，同时不会破坏所测定的蛋白质。此外，尚在使用的方法还有原子力显微镜(atomaticforce microscopy，AFM)对测定物进行表面扫描，激光捕获显微解剖技术(laser capture microdissection，LCM)以及荧光共振能量转换等，所有这些技术均是用来了解分析物中目标蛋白的结合变化情况的。目前的发展趋势是将各种微泵、微阀、微流路、微电泳技术运用到芯片上，以提高其分析检测的速度和灵敏度。综上所述图 18－3 显示蛋白质芯片检测的大概流程图。

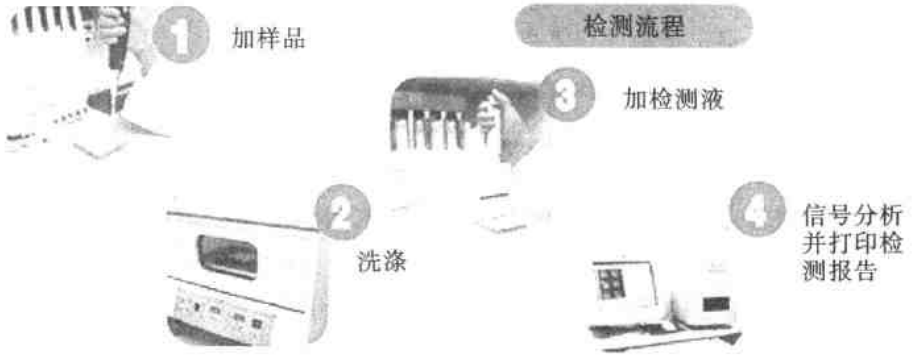


图 18－3　蛋白质芯片检测流程

(八)阅读结果,分析每一点的反应强度

例如 Lueking A 等人首先将人胚脑 cDNA 表达文库克隆至载体上,再用 IPTG(Isopropyl－β－D－thiogalactopymno－side)诱导表达包含 His6 表位的融合蛋白质。全过程均在 96 孔的微量滴定板上进行。接着将 96 孔的微量滴定板中的呈溶解状态的蛋白质点在 PVDF(polyvinylidene－difluoride)膜上。PVDF 膜具有极高的蛋白质结合力和机械强度。他们利用的是一种自动化转运印章技术可以进行三维定位,转运印章设计的点样密度上限是在一张 222 mm×222 mm 大小的膜上排 10 万个点,在 1 平方厘米上最多能点上 300 个样品。一个转运印章包含了 16 个点样头,其中 4 个点的是对照,另外 12 个点的是 6 种样品,一式两份。蛋白质点在 PVDF 膜上之后,蛋白质芯片就制成了。最后,作者在芯片上进行了普通的酶联免疫吸附检测,用一种能与含 His6 表位反应的单抗检测 PVDF 膜上排列的重组融合蛋白质。非特异性结合的抗体被洗掉,滤膜与荧光标记的二抗以及底物一起孵育。滤膜暴露于紫外线,阳性反应的点与阴性反应的点呈现不同的颜色,反应得到的图像经过 CCD(charge－coupled device)系统扫描分析,并由 AIDA 软件包处理。产生的信号是鲜明的、定位好并且可以检测 250 nmol 或者 10 pg 的点样的测试蛋白质(如 GAPDH)。蛋白质芯片的检测浓度下限是 10 fmol/μL。他们发现这种蛋白质芯片确实能够用于蛋白质表达和抗体筛选,并提供了一种高通量的配体－受体反应研究手段。作者认为,在利用 DNA 芯片进行基因表达谱分析的同时,可以同时利用蛋白质芯片进行蛋白质表达谱的分析,并进而分析多种蛋白质之间的相互作用。

四、蛋白质芯片的功能作用

蛋白质芯片通过标记可能存在反应性的蛋白质(如配体、酶等),使用特异性抗体或质谱测定,检测和筛选多种蛋白质之间的相互作用。

(一)抗原－抗体相互作用

De Wildt RMT 等人创建了一个抗体阵列,用来进行高通量的抗原－抗体反应筛选。首先将细菌文库的单菌落排列在第一张膜上,膜上的细菌菌落可以分泌蛋白产物(抗体)。第二张膜上印有抗原,能够与第一张膜上的抗体结合。将第一张膜压在第二张膜上,抗体就被第二张膜上的抗原粘附,于是第二张膜就具有了第一张抗体阵列。作者利用抗体阵列,与细胞裂解产物进行反应,筛选存在于细胞裂解物中的所有能与抗体阵列反应的抗原,研究细胞内的各种蛋白质表达水平。二维电泳是现在通用的分离复杂成分中蛋白质组的有效技术,但是过程比较

复杂，而且可重复性存在一定的问题。抗体阵列能够与细胞裂解物直接反应，比二维电泳方便快捷，得到的信息更多，只是灵敏度仍需进一步提高。

(二)蛋白质－蛋白质相互作用

这是蛋白质芯片最常用的研究目标。Thomas kodadek 报道了蛋白质功能芯片探测蛋白质－蛋白质之间相互作用。分别将 Protein G、P50 和 FKBP 上的 FRB(FKBP－rapamycin binding)区域点在化学处理后的玻片上，每种点四个点。然后用不同的荧光标记蛋白探测。如图 18－4 所示，A 板用能与 Protein G 结合的 BODIPY－FL－IgG 探测；B 板用能与 P50 结合的 Cy3－labeled－IkBa 探测；C 板用在雷帕霉素存在的情况下能与 FRB 结合的 Cy5－labeled－FKBP12 探测，其中，雷帕霉素存在于缓冲液中。D 板除了不加雷帕霉素外，其他与 C 板一样。E 板是在雷帕霉素存在条件下，同时探测 3 种标记蛋白。

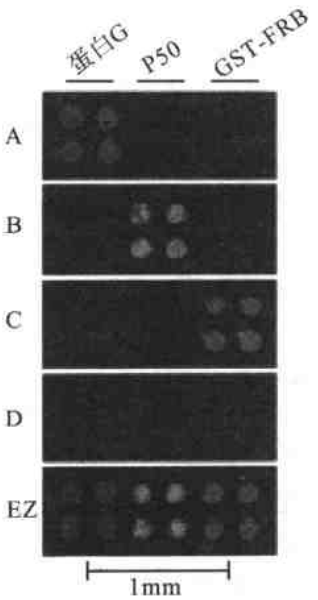


图 18－4　蛋白芯片探测蛋白质之间的相互作用

(三)酶－底物相互作用

Heng Zhu 和 Michael Snyder 报道了蛋白质微孔芯片在激酶阵列中的应用。将微孔芯片的表面用交联剂 GPTS(3－glycidoxypropyltrimenthoxysilane)激活并在 130 ℃纯化 1 h。蛋白质底物加入孔中，在湿度室中放在冰上放置 3 h。每孔用 1%的 BSA 封闭，用 TBS(tris－buffered saline)洗脱，然后加入激酶，$\gamma-^{33}P$－ATP 和缓冲液，接下来在 30 ℃孵育 30 min。洗净芯片，暴露于 MDP(molecular dynamics phosphoimager)，见图 18－5(a)。图 18－5(b)是用 poly(Glu－tyr)作底物得到的磷酸化结果放大图。

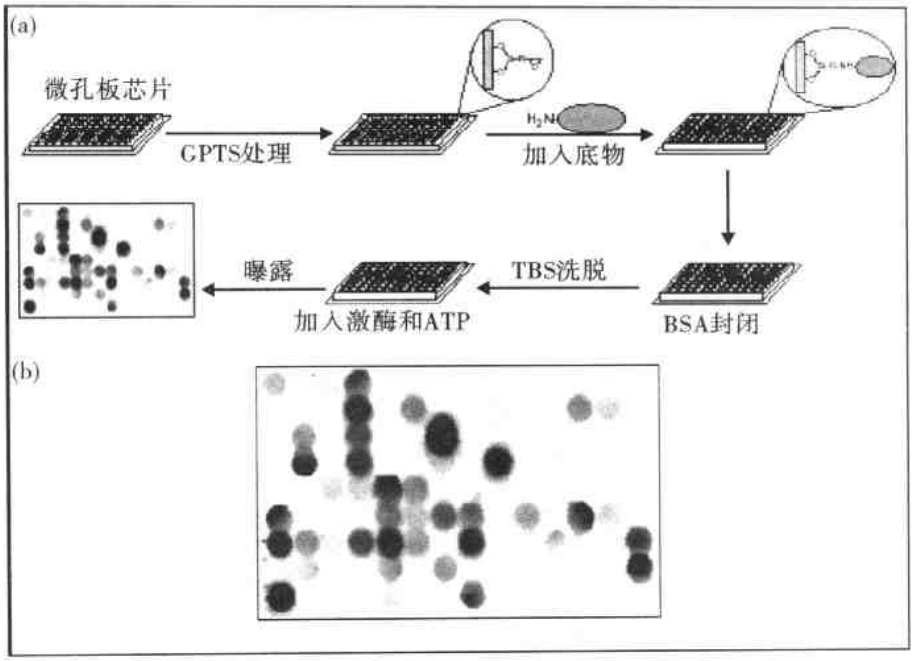


图 18-5　蛋白芯片探测酶-底物间的作用

(四)配体-受体相互作用

捕获和分析特异性标记蛋白质配体包被的表面现在已经有了,包括 SELDI 蛋白质芯片和多种 Blacore 芯片。表面细胞质基因组共振,此项 Blacore 应用的新进展对芯片模式的相互作用的研究贡献很大。SELDI 和 BIA-MS 技术将亲和捕获和高分辨率质谱分析与新的、有效定性蛋白质和相互作用的配体的手段结合了起来,并有诸如表位作图和受体-配体或蛋白质-药物相互作用研究的应用。

(五)高通量的蛋白质表达筛选

Buss WK 等人用类似的技术完成了高通量的蛋白质表达筛选。先是将 cDNA 文库克隆到原核表达载体,并用仪器将其克隆菌落排列在微量滴定板上,再转移到 PVDF 膜。接着用抗体与膜上的蛋白质反应,筛选反应阳性的克隆,作为第一轮筛选。第二轮筛选:小规模的蛋白质表达(指上轮的阳性克隆)在微量滴定板上进行,再用 SDS-PAGE、亲和层析、质谱分析表达产物的大小、产量、纯度、溶解度。对克隆的表达水平同时进行分析,以期发现最适于高通量结构分析与功能筛选的克隆。发现 66% 的克隆包含插入片段,并且插入方向正确,其中又有 64% 的克隆包含了完整的插入片段,能够编码一个完整的蛋白质。第三轮筛选:在微量滴定板上对蛋白质功能进行分析,测定了 96 个克隆的细菌裂解产物表达 GADPH 的活性。通过三轮筛选,作者建立了一个人类蛋白质表达克隆目录,为今后的结构与功能分析提供了基础。

(六)检测翻译后修饰

例如许多细胞功能通过可逆磷酸化调节,特定位点磷酸化位点及磷酸化程度的鉴定对于理解这些修饰如何调节受体活性至关重要。蛋白芯片与 MS 技术结合可快速、灵敏地用于翻译后修饰的测定。

耶鲁大学的 Micheal Snyder 等人构建了一组酵母蛋白质组芯片来筛选一些蛋白质的生化活性。将 5 800 个酵母开放阅读框克隆到酵母高拷贝表达载体上，使用可诱导的 GAL1 启动子在酵母中表达。采用 96 孔板，利用谷胱甘肽琼脂糖珠从酵母提取物中迅速纯化出 1 152 个样品。利用免疫印记分析随机的 60 个样品来监测纯化蛋白质的质量和数量。代表 5 800 种不同蛋白的 6 566 个蛋白样品通过商业购买的点样仪印到玻璃载玻片上。用抗 GST 的抗体来检测芯片，超过 93.5% 的蛋白样品有显著信号。在钙存在的条件下用生物素标记的钙调蛋白来检测酵母蛋白质，鉴定出六种已知的钙调蛋白标靶，以及一种结合 Cy－3 标记的抗生素蛋白链霉菌蛋白，说明蛋白质芯片有可能鉴定出许多转录后修饰蛋白。含有磷脂酰胆碱和 1% 生物素标记的 DHPE 的 5 类 PI(磷脂酰肌醇)脂质体和 1 类缺乏 PI 的脂质体用于检测蛋白质—磷脂相互作用。图 18－6 结果表明体内许多膜蛋白包括完整的膜蛋白和外周蛋白优先与特异的磷脂结合。

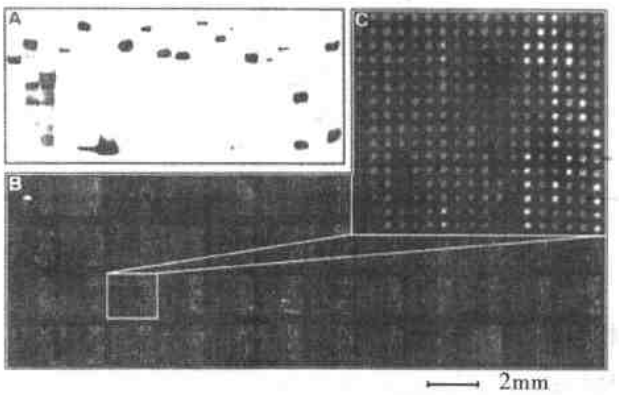


图 18－6　酵母蛋白芯片分析信号

五、蛋白质芯片的相关技术

(一)基因芯片

基因芯片与蛋白质芯片同属于生物芯片范畴，两者在构成、技术框架、基本原理等方面都有很多相似或相近之处，这是很明显的，所以有人称两者为“堂兄弟”。由于蛋白质芯片出现较晚，技术上也不如基因芯片成熟，所以很多学者认为蛋白质芯片是基因芯片的补充。实际上除了时间，某些技术(如蛋白质比 DNA 难合成，更难在固相支持物表面合成)等方面的差别外，两者在研究角度和发展理念上存在质的差别，这就决定了蛋白质芯片具有独立的研究品格。cDNA 芯片是从细胞中 mRNA 的角度来考虑的，其前提是细胞中 mRNA 的水平反映了蛋白质表达的水平。但事实并不完全如此，从 DNA→mRNA→蛋白质，存在三个层次的调控，即转录水平调控、翻译水平调控、翻译后水平调控。从 mRNA 角度考虑，实际上仅包括了转录水平调控，并不能全面代表蛋白质表达水平。实验也证明，组织中 mRNA 丰度与蛋白质丰度的相关性并不好，尤其对于低丰度蛋白质来说，相关性更差。更重要的是，蛋白质复杂的翻译后修饰、蛋白质的亚细胞定位或迁移、蛋白质－蛋白质相互作用等则几乎无法从 mRNA 水平来判断，而蛋白质才是生命活动的体现者和执行者，从这个意义上讲，蛋白质芯片更具有发展潜力。

(二)抗体芯片

抗体芯片(antibody microarray,抗体微阵列),是蛋白质芯片的一种,是检测生物样品中蛋白表达模式的新方法。这种新技术使得研究人员可以在一次实验中比较生物样品中成百上千的蛋白质的相对丰度,将极大促进蛋白质组目前的研究状况。之所以将抗体芯片单独提出,是因为其在蛋白质芯片家族中占有极其重要的地位,是其他类蛋白质芯片的技术基础,也是蛋白质芯片发展的原始平台。

有的研究者将抗细胞表面标志 CD 抗原的 CD 系列抗体蛋白固定在醛基玻片上形成微阵列,从样品细胞中抽提膜蛋白抗原并用 FITC(fluorescein isothiocyante)标记,然后与固定在芯片上的 CD 抗体阵列发生反应,扫描仪检测反应结果。发现所得结果与免疫印迹及免疫组织化学验证等方法一致,显示出其良好的适用性。

下面以 Clontech 公司第一代的抗体芯片 Ab Microarray 380(Cat. No. K1847 - 1)为例,简单说明抗体芯片的技术流程。该芯片包含固定在玻璃片基上的 378 种已知蛋白质的单克隆抗体,可以在一次简单实验中同时检测样品中的 378 种蛋白质的表达情况,并且可以在一张芯片上对两种样品的表达模式进行比较分析。

抗体芯片并不要求特殊的技能,只要一般常规的操作就可以完成以往极为复杂耗时的工作。整个操作流程包括:从 50 ~ 200 mg 组织或细胞、体液中进行蛋白质抽提→用 Cy5 和 Cy3 两种不同颜色的荧光分子分别标记两个样品→洗去多余的标记分子→与芯片杂交孵育→扫描分析结果。整个过程从样品制备到结果分析只要一天即可完成,你只要准备好样品、荧光染料、脱盐纯化柱(处理体液样品时用)和荧光扫描仪,其他的试剂全部由试剂盒提供(如图 18 - 7 所示)。

图 18 - 8 显示的是不同微阵列的技术、性质和功能比较。

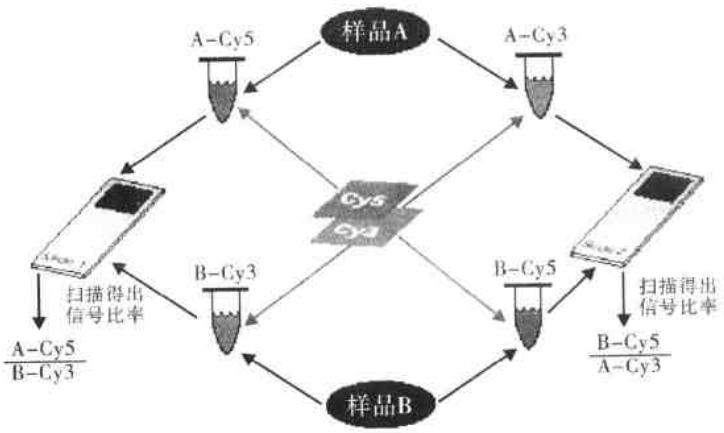


图 18 - 7

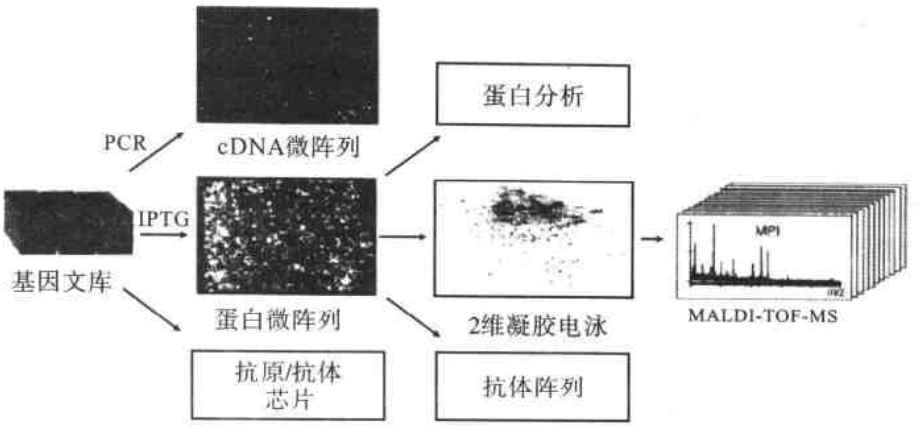


图 18-8 不同微阵列的技术比较

(三)其他相关技术

1.寡核苷酸芯片。一个凝胶光学或过硫酸盐诱导的共聚合技术已经被发展来制作聚丙乙烯凝胶垫上的寡聚核苷酸、DNA 和蛋白质微芯片三维阵列(从 10 μm × 1 μm 到 100 μm × 100 μm),其为一个憎水玻璃界面所分隔。

2.组织芯片。使用自动穿孔柱(0.6 mm 宽,3~4 mm 高),采取 1 000 个包埋在石蜡中的个体肿瘤活检组织,将它们排列到 45 mm × 20 mm 石蜡块中(可以从 Beecher Instruments, Silver Spring, MD, USA 购得)。连续切片后,肿瘤标本进行免疫组化、荧光原位杂交(FISH)和 RNA/RNA 原位杂交的平行分析。此系统能够在不到 2 h 内显微扫描含有 645 个标本的免疫组化阵列玻片。它对于许多处于不同疾病阶段的不同患者的同步分析,快速建立新的候选标记基因中有应用。

3.蛋白质光学芯片。此方法将高分辨率的生物光学显微成像技术和集成化多元蛋白质芯片技术相结合,发展形成了新型的并行、快速生物分子识别和检测技术。它不需要任何添加剂,对生物活性无影响,还可以进行生物分子相互作用的动力学检测,为一些尚无答案的生物医学理论问题提供解答。

其他还有微流体芯片,生物传感器等等。

(四)蛋白质芯片的技术问题

由于蛋白质具有多维立体结构,在多肽表面发生的蛋白质—蛋白质的相互作用分析比较困难,这是由于:

1.在微阵列制造时,必须维持蛋白的折叠性和立体构象,在蛋白质芯片制造过程中使用化学药品、加热或干燥等将影响蛋白质的折叠性,进而影响芯片质量。

2.折叠蛋白质之间存在相互作用,这种作用对线形序列有较强的依赖性,使得杂交验证反应的动力和定量分析复杂化。

3.制备高质量的芯片,还要得到荧光标记的蛋白质样品,这个技术还不成熟。

第三节 蛋白质芯片在医药学上的应用进展

蛋白质芯片在识别特定组织或细胞的蛋白质表达图谱、进行蛋白质水平的药物筛选、比较

相关蛋白间的特异性差异、揭示蛋白激酶的作用、测定血清中的小分子物质含量等方面均证实较现有技术更准确迅速。概括地讲，蛋白质芯片在临床研究中的作用大致分为三方面：

一、发现新的疾病标志物作为临床诊断的指标

利用蛋白质芯片对组织，细胞或体液中蛋白质表达的整体变化进行图谱分析，可获得蛋白质水平上的总体表达情况，通过对所谓表型指纹（phenomic fingerprint）的检测，即样品中与某种疾病或环境因素损伤可能相关的全部蛋白的含量变化状况进行检测，可以更可靠地开展疾病的诊断或筛查。尤其在临床医学方面，对某些疾病，特别是癌症及遗传性疾病的相关蛋白的识别上取得了一系列突破性进展。

（一）老年性痴呆

Davies H 等人利用蛋白质芯片和表面增强激光解吸附/离子化（surface enchanced laser desorption/ionization，SELDI）质谱来研究人培养细胞中 β 淀粉样变蛋白肽（amyloid－β peptide，Aβ）中的 39～43 个氨基酸残基，该片段被认为是人脑产生神经系统退行性改变的标志物，可导致基因突变并且能够毒害神经细胞的变异情况。制备蛋白质芯片时，将高选择性的抗－Aβ（抗体）共价偶联在芯片上，并将患者的组织细胞培养液和脑脊液滴加在芯片表面，与芯片反应，芯片上抗－Aβ 能够捕获液体中谱对芯片进行分析，显示靶蛋白的分子量和指纹图谱。早老性痴呆（又名阿尔茨海默病）患者的典型病理特征是老年斑与神经纤维缠结，其中老年斑主要由 Aβ 聚集而成。作为跨膜蛋白，淀粉样蛋白前体在水解时形成多种形式的 Aβ，其中以 Aβ（1～42 个氨基酸）最为危险，故研究 Aβ 的分型及各型的含量对于早老性痴呆的诊疗具有重要意义。这种检测方法具有相当高的敏感性和极微量的定量分析能力，所需样品中只要有几百个有相关改变的细胞就能检测出来，实验在 1 周内即可完成，对早期病变的快速检出具有重大的临床意义。比传统的检测技术先进得多。目前在此基础上正在探索制成相应抗体进行人群筛查，有相当大的发展潜力。

Frears ER 等人利用同样的技术分析胆固醇在 Aβ 的生物合成中的作用。

（二）癌症相关蛋白

蛋白质芯片在此领域进展最快。Watkins B 将乳腺癌、结肠癌、前列腺癌患者与非癌对照者的血清或排泄物样品通过蛋白质芯片进行了研究。分别用 5 种蛋白质芯片来分析，这些芯片包括含有反向结合的疏水表面、结合阳离子、结合阴离子、结合与铜有亲和力的蛋白、结合与镍有亲和力的蛋白几种类型，根据特定条件下检测出的不同图谱寻找在癌症患者中存在而在正常人中没有的特殊差异蛋白，然后从这些蛋白里选择出 1～2 个来进一步在其他癌症患者及健康人中加以验证。此后再在不同类型、不同阶段的恶性病变及良性病变中进行比较性研究，以确定这一蛋白的临床意义。现在经 SELDI－TOF 分析发现的癌症相关蛋白有以下几种。

1. 乳腺癌特异蛋白。乳腺癌患者的血清样品中有一个 28.3 kD 的蛋白，而且它仅存在乳腺癌患者的样品中。由此，研究者设计出了 NMP66 试剂盒，用其对可疑患者进行检测，结果发现这一蛋白在乳腺癌的早期诊断（无腋下淋巴结转移）及晚期转移性乳腺癌的诊断上特异性达到了 100%，在排除非恶性肿瘤个体上的特异性达到 96%，远远高于临床应用的其他检测手段（如乳房 X 线造影及红外线检查），也比目前 FDA 批准的判断乳腺癌预后相关的血清蛋白 CA15－3 和 CA27－29 的敏感性和特异性好。而且该蛋白在 DCIS（复发性非浸润性恶性肿瘤）的诊断敏感性上亦达到 80%，可作为预测 DCIS 术后复发及治疗的重要监测指标，有望被作为

一个公认的、比以往发现的其他各种标记物更好的乳腺癌诊断标志。

2. 结肠癌特异蛋白。结肠癌患者的样品中发现有一个 13.8 kD 的特殊相关蛋白，存在于结肠癌及其癌前病变患者的血清中，而非癌或克罗恩病以及溃疡性结肠炎等良性肠紊乱病的患者中该蛋白缺失。因此通过检测血中是否含有该蛋白存在可确定肠病的良、恶性质并对结肠癌做出早期筛查。肠道憩室炎患者的血清里也发现了这一特异蛋白，其原因值得进一步深入研究。

3. 前列腺癌特异蛋白：前列腺癌的样品中发现一个 50.8 kD 的蛋白存在于全部患者的血清或尿液中，正常健康对照者均为阴性，而且通过血清前列腺特异性抗原(PSA)检测未被检测出的前列腺癌患者中也能检测出该蛋白质。因此 Paweletz 等认为用蛋白质芯片的方法能够检测出当前常用的 PSA、前列腺的特异肽、特异性酸性磷酸酶、特异膜抗原、癌细胞裂解物的检查方法尚不能识别的小分子癌相关蛋白。在正常人、癌前期及癌症不同时期的患者身上，利用蛋白质芯片技术，仅用约含 500～1 000 个左右组织细胞的样品就迅速将几种前列腺癌的相关蛋白寻找出来，敏感性特异性均优于目前一系列其他的检测方法。Ciphergen biosystems 公司利用蛋白质芯片检测了来自于健康人和前列腺癌症患者的血清样品，在短短 3 天之内发现了 6 种潜在的前列腺癌的生物学标志。如果用过去的方法，也许要花费数月到数年的时间。

4. 早期识别卵巢癌的初步研究：Brown 报道了应用 LCM 蛋白质芯片技术评估卵巢组织分离细胞中生殖蛋白的研究结果。在 3 种卵巢肿瘤组织中发现了 5～8 个具有不同分子量的特殊肽段，这为今后早期发现并诊断出这些疾病迈出了可喜的一步。

5. 尿液中移行肿瘤细胞的检出：Vlahou 等将含肿瘤移行细胞的尿液进行了 SELDI－TOF 蛋白质芯片分析，共测定了含有肿瘤移行细胞的患者尿液、患其他泌尿系统疾病患者的尿液及健康人的尿液 94 份。发现肿瘤患者的尿样中有多种蛋白质的改变，其中包括 5 种异常的新蛋白和 7 种异常的蛋白组合，分子量从 3.3～13.3 kD 不等，用这种方法判断膀胱中有无移行的肿瘤细胞的敏感性可达 78%，而传统的膀胱冲洗后尿液的细胞学检查敏感性仅仅才 33%。

(三)其他

Howard 等利用该方法对乳酸杆菌中的胶原结合蛋白进行识别。Thulasiraman 等关于 Yersinia 黑死病病毒特点的研究，Li 等通过动物耳组织对软组织受损后促进修复基因表达的筛选中也均使用了蛋白质芯片的研究技术。在实验室里用蛋白质芯片筛查特异性抗体和测定某些特定的基因表达异常则比较普遍。赵芷音等将抗原性强的 C、NS3、NS4、NS5 片段用芯片点样仪定量点加在醛基化玻片上来制作蛋白芯片。用进口的 HCV 检测试剂盒(ELISA)筛选出 65 例抗－HCV 阳性标本，阳性标本用芯片检测不同功能区抗体，并测定 HCV－RNA；同时用芯片检测 24 例抗－HCV 阴性的正常献血员血清。结果发现 24 例抗－HCV 阴性标本用芯片检测均阴性；65 例抗－HCV 阳性标本经芯片检测均有 1～4 种抗体阳性，其中抗－C、抗－NS3、抗－NS4、抗－NS5 阳性率最高，占 33.8%；抗－C、抗－NS3、抗 NS5、抗 NS4 总的检出率分别为 98.5%、89.2%、52.3%和 50.8%；未发现单独含抗－NS3、抗－NS4 的血清；抗－C、抗 NS3、抗－NS4、抗－NS5 阳性血清中 HCV－RNA 检出率最高，占 77.3%；HCV－RNA 阳性血清中抗－NS5 阳性率最高，占 61.8%，说明抗－NS5 与 HCV－RNA 最有相关性($P<0.05$)。结论：蛋白质芯片显示较高的敏感性与特异性；抗－C、抗－NS3 在诊断抗－HCV 都很重要，抗－NS4 及抗－NS5 有诊断的互补作用，而抗－NS5 在一定程度上能反映 HCV 的复制情况。此外，目前国内已经有成熟的多肿瘤检测系统问世(图 18－9)。

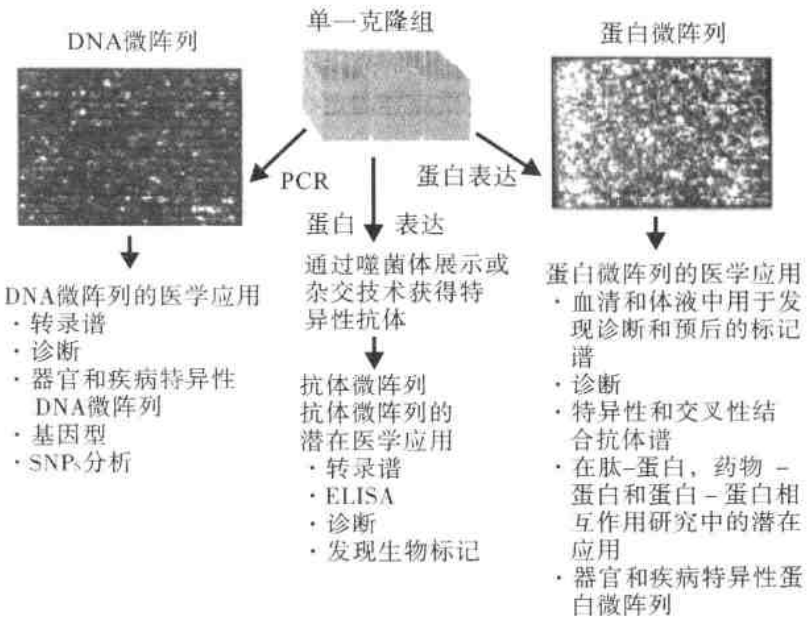


图 18-9　多肿瘤检测系统

二、对疾病病理分子机制的研究

蛋白质芯片能剖析疾病发生的分子基础，致使临床医生能够从疾病的成因而不是以其症状进行诊断，并针对这种分子变化予以治疗措施。一个常用的例子是在明确 *Her*2 癌基因过度表达是乳腺癌发病的分子机制后，便可用治疗性抗体 Herceptin 加以治疗。蛋白质芯片对分子机制的探讨主要围绕病变细胞接受和传递环境刺激的分子信号途径以及在生物个体的生理和病理状态下这些刺激对蛋白质表达的调控途径，是目前肿瘤细胞信号转导应用研究中进展较快的领域。根据细胞生命活动类型，信号转导形式可分为增殖相关通路、凋亡相关通路等。Paweletz 等首先将反向蛋白质微阵列(reverse phase protein microarry，RPPM)应用于恶性肿瘤侵袭边缘 prosurvival 通路活化的研究。常用的蛋白质阵列是将探针固定，而 RPPM 是将来源于病变不同时期的细胞蛋白质固定。这种技术的高敏感性、精确度和线性关系可辨别信号蛋白的磷酸化情况。通过构建前列腺正常上皮、皮内肿物和侵袭性癌的蛋白质阵列，可以分析 prosurvival 关键点蛋白质状态，结果显示前列腺癌的侵袭性与 Akt 磷酸化增加、凋亡通路抑制以及 ERK 磷酸化降低有关；其中 Akt 的改变在正常上皮向 PIN 过渡期尤为显著，伴有下游凋亡通路抑制。以胞浆激酶类型的不同存在 MAPK 激酶通路、PKC 激酶通路等。新西兰奥克兰大学组建了人脑海马区的蛋白质组图库，鉴定了在精神分裂症中不正常表达的 18 种蛋白质，他们正在研究这些蛋白质在疾病病理过程中的作用。

三、筛选药物靶，发现新药物，研究药理学

在新药研究中蛋白质芯片方法主要应用于受体或标志物的鉴定，以及受体的评价及毒理学，药物代谢产物的研究。通过对比健康状态与疾病状态的细胞或组织的蛋白质表达差异用于药物研究或药物受体的研究，或药物治疗前后蛋白质表达状况的分析，以评价药物类似物的

结构与活性关系，寻找高活性的药物。蛋白质芯片还用于药物毒性的研究，正常组织或已知毒性试剂处理的组织其蛋白质表达情况的比较，可以表明药物的毒性及活性。耶鲁大学的研究小组报道的真核生物水平的蛋白质微阵列，他们克隆了 5 800 个开放阅读框，进行酵母蛋白质组的过量表达及蛋白纯化，并将这些蛋白质点样固定在载玻片上，制作了蛋白质芯片，可应用于筛选蛋白药物相互作用（见上文）。

Gan Z 等人利用生物素酰基化酶抑制吸附检测（biotinlated enzyme inhibitorsorbent assay，BEISA）技术来寻找特定酶的抑制剂，而这些抑制剂往往可以成为药物。他们先将多种抑制剂固定在膜上，再用生物素标记的酶和未标记的抑制剂与膜共同孵化，洗去未结合上的酶、抑制剂，接着加入链霉素抗生素蛋白－碱性磷酸酶，与膜共育，最后加入碱性磷酸酶的作用底物，产生颜色反应。洗脱之后，观察各个位点的颜色，就可以知道哪些抑制剂能够与酶结合，结合的强度如何。由于在一张膜上可以排列上千种抑制剂，因此这也是一种高通量的蛋白质功能筛选体系。

G 蛋白偶联受体一直是重要的药物作用靶点。受体受到刺激之后，会导致细胞内钙离子的浓度变化。Coward P 等人将 4 种 G 蛋白偶联受体制成各种浓度梯度，并固定在微量滴定板上，再分别加入 31 种兴奋剂。通过测量钙离子的浓度变化，来判断兴奋剂与受体之间的关系。作者已经有能力可以同时点上千种受体，并且交叉反应污染较少，为研究复杂的配体受体反应提供了一种高通量的检测方法。他们使用了一种自动化的信号检测装置荧光计图像板阅读器（fluorometric imaging plate reader，FLIPR），平行、实时检测微量滴定板中的钙离子的浓度变化。

蛋白质芯片在肿瘤和其他疾病中可用于药物靶向的研究与新药的研制，同时也可用于自身免疫性疾病的诊断。德国学者 Leuking 用人胚脑 hExl 文库中的 92 个表达克隆产物及 4 例对照标本制成的蛋白质芯片与单克隆抗体 RCS－His 孵育，检测 RCS－His6 标记融合蛋白在每一克隆的表达，以往的研究曾证实 hExl 表达此蛋白。在 92 个表达克隆中有 54 个克隆的插入序列与 GenBank 登记的人类基因序列一致。经序列分析编码 RCS－His6 标记融合蛋白的读框发现，34 个克隆具有正确的读框，20 个克隆是不正确的读框序列，表达产物鉴定后假阴性率在 2% 以下，假阳性率在 11% 以下，而蛋白质的原位膜杂交假阳性率 37%，Leuking 等人仍用上述蛋白质芯片分别与人类三磷酸甘油醛脱氢酶（GAPDH）、热休克蛋白 90α 的羧基端片段（HSP90α）和 α 微管蛋白相应的单克隆抗体反应，以观察单克隆抗体的特异性。抗 GAPDH 抗体特异地识别了其靶抗原，无任何交叉反应，而抗 HSP90α 抗体、抗 α 微管蛋白抗体分别与另外 2 个和 1 个克隆发生交叉反应，其中 2 个克隆所代表的基因是未知的。这一方面说明蛋白质芯片可用于检测已知抗体的特异性；另一方面根据抗体的交叉反应可以发现蛋白质相互作用的新途径，特别适用于配体 A 受体及药物作用机制的研究。

我国具有丰富的中药资源，并拥有成熟完整的理论和应用方式，值得深入开发研究。在中药品种优化及鉴定中，蛋白芯片技术与中药植物化学结合，为其开辟了一条新路，可广泛用于分析药用植物种属群体及个体的鉴别。

第四节　蛋白质芯片的发展展望

蛋白质芯片的问世，提供了一个有效的分析平台，在研究蛋白质的性质、特征、辨识、功能及生物标记等方面架起快速准确的桥梁，成为一种极具发展前途的新型研究工具，极大的推动了蛋白质组学、蛋白质生物信息学的迅速发展。蛋白质芯片作为研究蛋白质组学的方法不仅需要应用已建立的技术诸如高性能的微阵列、大规模的表达和高分辨率的二维电泳技术，更需

要新发展的自动化技术。这些包括自动化二维电泳、凝胶成像、点识别、蛋白质裂解后的点切除、样品纯化和质谱分析以及闭环数据获取和处理的软件。一个高通量 MALDI－TOF－MS 技术近来已经被发展进行有力的、快速和高精度的蛋白质控制。高通量技术的发展对于这样的蛋白质分析是需要的，因为大数量的蛋白质参与机体的自稳态，即使是最简单的有机体也是一样的。

目前，以下几个领域是蛋白质芯片的研究趋势：

1.寻找更好的探针固定技术。

2.加速样品处理和标识研究。

3.研究新的仪器和方法。不断涌现的新技术可以加速蛋白质芯片的发展，如将寡核苷酸阵列转化为多肽和蛋白质阵列的核糖体展示技术，RNA－多肽融合技术，寡核苷酸标签配体技术和大规模的利用基因替换与 RNA 抑制进行的基因敲除技术。在检测技术方面，荧光共振能量转换和原子力显微镜都可以用于蛋白质芯片的研究。具有网上检测的高通量自动化芯片杂交设备，加上整合的图像和数据分析工具的改进将加快 DNA 和蛋白质芯片输出的速度和质量。

4.开发高度集成化生产的制备系统等。

5.改进蛋白质表达系统。近来研制了一个在 Pichi pastoris 和 *E. coli* 中双蛋白质表达系统，它结合了两个系统的优点，诸如可溶性蛋白质表达的改善（A. Lueking 等，未发表），但不需亚克隆。这样表达的可溶性蛋白质对于许多应用是需要的，比如应用于治疗、体外相互作用研究、生成抗体识别结构表位以及结晶和核磁共振（NMR）研究。

6.蛋白质数据库的完善。高分辨率、高敏感性和高通量的分离和分离后的鉴定技术，结合准确、全面的数据库技术，使蛋白质芯片技术用于分子生物学研究卓有成效，必将会促进未来生命科学的整体发展。但是，蛋白质芯片研究还处于技术发展和创新期，蛋白质组信息学一个主要挑战是整合能力不足，今后的研究将是整合整个蛋白质组信息，如从二维电泳凝胶、质谱和序列，到代谢与功能的数据，以及蛋白质结构和相互作用，但也应联系到相关基因序列和表达数据，从而建立一个使各种数据与分析仪器一体化的信息学操作平台。

蛋白质芯片发展的下一步是能够对更复杂，更系统的蛋白质组，也就是说在一张芯片上排列至少 5 000～10 000 个蛋白质。芯片上的探针主要来自重组抗体、噬菌体展示文库，这里面可以储备上亿个的抗体片段。制作蛋白质芯片的关键是具备高容量的抗体文库，也许在不久的将来，我们就可以利用蛋白质芯片来分析各种蛋白质组，但是目前的工作离这一目标还有相当远的距离。另一方面，蛋白质芯片之所以比 DNA 芯片发展慢，主要还是因为蛋白质本身的变化太多，如何保证检测结果的可重复性，如何降低成本，这些都是需要科学家们进一步加深对蛋白质结构和功能的理解，需要其他相关辅助技术的成熟。用芯片技术取代传统的对单个基因和蛋白质进行分析的技术，目前还有较大困难，但是，蛋白质芯片提供了一种高通量功能分析技术，作为一种分析复杂生物系统的思路，是值得我们高度重视的。在应用方面蛋白质芯片似乎是蛋白质组学的新的和多功能工具。人们认为蛋白质芯片将使高通量筛选广泛种类的分子间相互作用成为可能。因为这是一个进步非常快的领域，期望制作出敏感性特异性更强且固定有多种大量活性蛋白的新型芯片来，对诊断、药物筛选、发育分析和蛋白质组学的研究做出真正贡献，有人甚至提议用蛋白质芯片组装成的纳米机器人，通过血管送入人体去侦察疾病、携带 DNA 去更换或修复有缺陷的基因片段。也可以携带纳米药物的芯片送入人体内，在外部加以导向，使药物集中到患处，更理想的提高药物疗效。还有学者在研究利用蛋白质芯片

组装成生物计算机。毋庸质疑，这项技术在新世纪里不仅对认识蛋白质组与人类健康错综复杂的关系，对疾病的早期诊断和疗效监测等会产生巨大的推动作用，而且在其他相关领域如环境保护、食品卫生、生物工程、工业制药等方面也将具有广阔的发展前景。

各学科之间的交叉进而获得技术上的突破或融合，仍然是打破蛋白质组/蛋白质研究瓶颈的关键，在可能的技术到位之前，蛋白质点阵(芯片)无疑仍将有一段艰难的路要走，《科学》杂志说2000年的突破在于对基因与表达调节、蛋白质结构与功能新领域的探索，尤其是基因和蛋白质点阵/芯片技术突破的奇迹，这个论断被证明是正确的，按照此思路，可以大胆预测，整个21世纪生命科学的突破必将得益于蛋白质芯片技术的不断完善。

(吴良志　韩　玲)

参考文献

1. Young RA, Davis RW. Efficient isolation of genes by using antibody probes. Proc Natl Acad Sci U S A, 1983, 80(5): 1194
2. Hagen FS, Gray CL, Kuijper JL. Assaying the quality of cDNA libraries. Biotechniques, 1988, 6(4):340
3. Kohler SW, Provost GS, Kretz PL, et al. Development of a short-term, in vivo mutagenesis assay: the effects of methylation on the recovery of a lambda phage shuttle vector from transgenic mice. Nucleic Acids Res, 1990, 18(10): 3007
4. Kretz PL, Kohler SW, Short JM. Identification and characterization of a gene responsible for inhibiting propagation of methylated DNA sequences in merA merB1 Escherichia coli strains. J Bacteriol, 1991, 173(15): 4707
5. Lisitsyn N, Lisitsyn N, Wigler M, Cloning the differences between two complex genomes. Science, 1993, 259: 946
6. Sargent TD, Dawid IB. Differential gene expression in the gastrula of Xenopus laevis. Science, 1983, 222:135
7. Wiesner RJ, Ehmke H, Faulhaber J, et al. Dissociation of left ventricular hypertrophy, β-myosin heavy chain expression, and myosin isoform switch in rat after ascending aortic stenosis. J Mol Cell Cardiol, 1997, 95:1253
8. Diatchenko L, Lau YFC, Campbell AP, et al. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci USA, 1996, 93: 6025
9. Zhu F, Yan W, Zhao ZL, et al. Improved PCR-based subtractive hybridization strategy for cloning differentially expressed genes. BioTechniques, 2000, 29:310
10. Velculescu VE, Zhang L, Vogelstein B, et al. Serial analysis of gene expression. Science, 1995, 270:484
11. Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the PCR. Science, 1992, 257:967
12. Zhang H, Zhang R, Liang P. Differential screening of gene expression difference enriched by differential display. Nucleic Acids Research, 1996, 24(12): 2454
13. Trask BJ. Fluorescence in situ hybridization: application in cytogenetics and gene mapping. Trends Genet, 1991, 7: 149
14. Yokota H, van den Engh G, Mostert M, et al. Treatment of cells with alkaline borate buffer extends the capability of interphase FISH mapping. Genomics, 1995, 25(2): 485
15. Ekong R, Wolfe J. Advances in fluorescene in situ hybridization. Curr Opin Biotech, 1998, 9: 19
16. Cox DR. Radiation hybrid mapping. Cytogenet Cell Genet, 1992, 59:80
17. Boehnke M, Lange K, Cox DR. Statistical Methods for Multiple Radiation Hybrid Mapping. Am J Genet, 1991, 49: 1174
18. Bishop DT, Crockford GP. Comparisons of radiation hybrid mapping and linkage mapping. Cytogenet Cell Genet, 1992, 59:93
19. Stringham HM, Boehnke M, Lange K. Point and interval estimates of marker location in radiation hybrid mapping Am J Hum Genet, 1999, 65 (2):545
20. Schultz J, Doerks T, Ponting CP, et al. More than 1 000 putative new human signalling proteins revealed by EST data mining. Nat Genet, 2000, 25(2):201
21. Cohen BA, Mitra RD, Hughes JD, et al. A computational analysis of whole-genome expression data reveals chromo-

somal domains of gene expression. Nat Genet, 2000, 26(2):183

22. Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A, 2002,99(6):3740

23. Wang Y, Newton DC, Robb GB, et al. RNA diversity has profound effects on the translation of neuronal nitric oxide synthase. Proc Natl Acad Sci U S A, 1999,96(21):12150

24. Fields S, Song O. A novel genetic system to detect protein – protein interactions. Nature, 1989, 340:245

25. F.奥斯伯, R.布伦特, R.E.金斯顿,等. 颜子颖, 王海林 译. 精编分子生物学实验指南. 北京:科学出版社, 1998

26 吴乃虎. 基因工程原理. 第2版. 北京:科学出版社, 1998

27. Bartel PL, Chien CT, Sternglanz R, et al. Elimination of false positives that arise in using the two – hybrid system. Bio Techniques, 1993(b), 14:920

28. Agrawal S. Antisense oligonucleotides: towards clinical trials. Trends Biotechnol, 1996, 14:376

29. Blatter EE, Ebright YW, Ebright RH, et al. Identification of an amino acid – base contact in the GCN4 – DNA complex by bromouracilmediated photocrosslinking. Nature, 1992, 359:650

30. Bless NM, Smith D, Charlton J, et al. Protective effects of an aptamer inhibitor of nertrophil elastase in lung inflammatory injury. Curr Bio, 1997, 7(11):877

31. Bridonneay P, Chang Y, O'Connell D, et al. High – affinity aptamers selectively inhibit human nonpancreatic secretory phospholipase A2 (hnps – PLA2). J Med Chem, 1998, 41:778

32. Boulianne GL, Hozumi N, Shulman M, et al. Production of functional chimeric mouse/human antibody. Nature, 1984, 312:643

33. Charlton J, Sennello J, Smith D. In vivo imaging if inflammation using an aoptamer inhibitor of human neutrophil elastase. Chem Bio, 1997, 4:809

34. Ciesiolka J, Illangasekare M, Majerfeld, et al. Affinity selection – amplification from randomized ribooligonuclettide pools. Methods in Enzymology, 1996, 267:315

35. Davis KA, Lin Y, Abrams B, et al. Staining of cell surface human CD4 with 2 – F – pyrimidine – containing RNA aoptamers for flow cytometry. Nucl Acids Res, 1998, 26:3915

36. Eaton BE, Gold L, Hicke BJ, et al. Post – SELEX combinatorial optimization pf aptamers. Bioorg Med Chem, 1997, 5:1087

37. Ellington AD, Szostsk JW. Selection in vitro of single – stranded DNA molecules that fold into specific ligand – binding structure. Nature, 1992, 355: 850

38. Gold L, Brown D, He U, et al, From oligonucleotide shapes to genomic SELEX: novel biological regulation loops. Proc Natl Acad Sci USA, 1997, 94:58

39. Guckion KM, Krugh TR, Kool ET, et al. Efficient replication between non – hydrogen – bonded nucleotide shape analogues. Nat Struct Bio, 1998, 5:950

40. Hicke BJ, Willis MC, Koch TH, et al. Telomeric protein – DNA point contacts identified by photo – cross – linking using 5 – bromodeoxyuridine. Biochemistry, 1994, 33:3364

41. Jellinek D, Louis SG, Carol B, et al. Inhibition of receptor binging by high – affinity RNA ligands to vascular endothelial growth factor. Biochemistry, 1994, 33:10450

42. Jenison RD, Gill SC, Pardi A, et al. High – resolution molecular discrimination by RNA. Sciences, 1994, 263:1425

43. Klausner RN, Rouault TA, Hardord JB. Regulating the fate of mRNA: the control of cellular iron metabolism. Cell, 1993, 72:19

44. Kraus E, James W, Barclay AN, et al. Novel RNA ligands able to bind CD4 antigen and inhibit CD4 + T lymphocyte function. J Immunol, 1998, 160:5209

45. Lieber A, Strauss M . Selection of efficient cleavage sites in target RNA by using a ribozyme expression library. Mol

Cell Bio, 1995, 15: 540

46. Liu J, Sodeoka M, Lane WS, et al. Evidence for a non－α－helical DNA－binding motif in the Rel homology region, Proc Natl Acad Sci USA, 1994, 91: 908

47. Meidenheimer KM, Meisenheimer PL, Willis MC, et al. High yield photocrosslinking of a 5－iodocytidine(IC) substituted RNA to its associated protein. Nucleic Acids Res, 1996, 24: 981

48. Pan T. Novel RNA substrates for the ribozyme from Bacillus subtilis ribonucleoase P identified by in vitro selection. *Biochemistry*, 1995, 34: 8458

49. Pieken WA, Olsen DB, Benseler F, et al. Kinetic Characterization of ribonuclease－resistant 2′－modified hammerhead ribozymes. Science, 1991, 253: 314

50. Prudent JR, Uno T, Schultz P. Expanding the scope of RNA catalysis. Science, 1994, 264: 1924

51. Roberts Rw, Szostak Jw. RNA－peptide fusions for the in vitro selection of peptides and proteins. Proc Natl Acad Sci USA, 1997, 94: 12297

52. Scott JK, Smith GP. Searching for peptide ligands with an epitope library. Science, 1990, 249: 386

53. Singer BS, Shtaland T, Brown D, et al. Libraries for genomic SELEX. Nucleic Acids Res, 1997, 25: 781

54. Tasset DM, Kubik MF, Steiner W. Olignucleotide inhihitors of human thrombin that bind distinct epitopes. J Mol Biol, 1997, 272: 688

55. Tuerk C, MacDougal S, Gold L. RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. Proc Natl Acad Sci USA, 1992, 89: 6988

56. Van Ness J, Laem mLi UK, Pettijohm DE, et al. Immunization in vitro and production of monoclonal antibodies specific to insoluble and weaky immunogeric proteins. Proc Natl Acad Sci USA, 1984, 81: 7897

57. Vaughan T, Osbourn JK, Jempest FR, et al. Human antibodies by design. Nat Biotechnol, 1998, 16: 535

58. Weigand TW, Williams PB, Dreskin SC, et al. High－affinity oligonucleotide ligands to human IgE inhibit binding to Fee receptor I. J Immunol, 1996, 157: 231

59. Willis MC, Hicke BJ, Uhlenbeck OC, et al. Photocroslining of 5－iodouracil－substituted RNA and DNA to proteins. Science, 1993, 262: 1255

60. Müller U. Ten year of gene targeting: targeted mouse mutants, from vector design to phenotype analysis. Mechanisms of Development, 1999, 82: 3

61. 杨晓，黄培堂，黄翠芬. 在小鼠进行基因打靶的研究进展. 科学通报，2000，45(15)：1584

62. Fuchs E, Segre JA. Stem cells: A new lease on life. Cell, 2000, 100: 143

63. McLaren A, Biggers JD. Successful development and birth of mice cultivated in vitro as early embryos. Nature, 1958, 182: 877

64. Stevens LC. The development of transplantable teratocarcinomas from intratesticular grafts of pre－ and post－implantation mouse embryos. Dev Biol, 1970, 21: 364

65. Papaioannou VE, McBurney MW, Gardner RL, et al. Fate of teratocarcinoma cells injected into early mouse embryos. Nature, 1975, 258: 70

66. Bradley A, Evans M, Kaufman MH, et al. Formation of germ－line chimaeras from embryo－derived teratocarcinoma cell lines. Nature, 1984, 309: 255

67. Thomas JA, Itskovitz－Eldor J, Shapiro SS, et al. Embryonic stem cell lines derived from human blastocysts. Science, 1998, 282: 1145

68. Reubinoff BE, Pera MF, Fong C, et al: Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. Nature Biotechnology, 2000, 18: 399

69. Doetschman T, Gregg RG, Maeda N, et al. Targeted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature, 1987, 330: 57657

70. Mansour SL, Thomas KR, Capecchi MR. Disruption of the proto－oncogene int－2 in mouse embryo－derived *stem*

cells: a general strategy for targeting mutations to non-selectable genes. Nature, 1988, 336:348

71. Askew GR, Doetschman T, Lingrel JB. Site-directed point mutations in embryonic stem cells: a gene-targeting tag-and-exchange strategy. Mol Cell Biol, 1993, 13:4115

72. Bedell MA, Jenkins NA, Copeland NG. Mouse models of human disease. Part I. Techniques and resources for genetic analysis in mice. Genes Dev, 1997, 11:1

73. Bedell MA, Jenkins NA, Copeland NG. Mouse models of human disease. Part II. Recent progress and future directions. Genes Dev, 1997, 11:11

74. Colvin JS, Bohne BA, Harding GW, et al. Skeletal overgrowth and deafness in mice lacking fibroblast growth factor receptor 3. Nature Genet, 1996, 12:390

75. Chen L, Monsonego EO, Yang X, et al. Gly369Cys mutation in mouse FGFR3 causes achondroplasia by affecting both chondrogenesis and osteogenesis. The Journal of Clinical Investigation, 1999, 104: 1517

76. Jacks T. Lessons from the p53 mutant mouse. J Cancer Res Clin Oncol, 1996, 122:319

77. Hakem R, de la Pompa JL, Sirard C, et al. The tumor suppressor gene Brca1 is required for embryonic cellular proliferation in the mouse. Cell, 1996, 85:1009

78. Xu X, Wagner KU, Larson D, et al. Conditional mutation of Brca1 in mammary epithelial cells results in blunted ductul morphogenesis and tumor formation. Nat Ggenet, 1999, 22(1):37

79. Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning, A laboratory manual. 2nd ed. Cold Spring Harbor Laboratory Press, 1989

80. Charreau B, Tesson L, Soulou JP, et al. Transgenesis in rats: technical aspects and models. Transgenic Res, 1996, 5(4): 223

81. Fujiwara Y, Miwa M, Takahashi R, et al. High-level expressing YAC vector for transgenic animal bioreactors. Mol Reprod Dev, 1999, 52(4):414

82. Brenin DK, Talamonti MS, Iannaccone PM. Transgenic technology: an overview of approaches useful in surgical research. Surg Oncol, 1997, 6(2):99

83. Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature, 2001, 409(6818):363

84. Brummelkamp TR, Bernards R, Agami R. A system for stable expression of short interfering RNAs in mammalian cells. Science, 2002, 296(5567):550

85. Elbashir SM, Lendeckel W, Tuschl T. RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev, 2001, 15(2):188

86. Nishikura K. A short primer on RNAi: RNA-directed RNA polymerase acts as a key catalyst. Cell, 2001, 107(4): 415

87. Paddison PJ, Caudy AA, Hannon GJ. Stable suppression of gene expression by RNAi in mammalian cells. Proc Natl Acad Sci USA, 2002, 99(3):1443

88. Tuschl T. RNA interference and small interfering RNAs. Chembiochem, 2001, 2(4):239

89. Marshall A, Hodgson J. DNA chips: An array of possibilities. Nature Biotechnology, 1998, 16: 2731

90. Schena M. DNA Microarrays, A Prectical Approach. OXFORD University Press Inc., New York, 1999, 2001 (three times reprimted).

91. Cheng J, Shoffner MA, Hvichia GE, et al. Chip PCR. II. Investigation of different PCR amplification systems in microbabricated silicon-glass chips. Nucleic Acids Res, 1996, 24(2): 380

92. Lueking A, Horn M, Eickhoff H, et al. Protein microarrays for gene expression and antibody screening. Anal Biochem, 1999, 270(1): 103

93. 马立人,蒋中华. 生物芯片.北京:化学工业出版社,2000

94. Yershov K, Barsky V, Belgovskiy A, et al. DNA analysis and diagnostics on oligonucleotide chips. Proc Natl Acad Sci

USA, 1996, 93:4913

95. Lockhart DJ, Dong H, Byrne MC, et al. Expression monitoring by hybridization to high - density oligonucleotide arrays. Nature Biotechnology, 1996, 14: 1675

96. Schena M, Shalon D, Davis RW, et al. Qantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, 1995, 270: 467

97. Shalon D, Smith JS, Brown PO. A DNA microarray system for analyzing complex DNA samples using two - color fluorescent probe hybridization. Genome Res, 1996, 6: 639

98. Dubiley S, Kirillov E, Lysov Y, et al. Fractionation, phosphorylation and ligation on oligonucleotide microchips to enhance sequencing by hybridization. Nucleic Acids Res, 1997, 25(12): 2259

99. Kozal MJ, Shah N, Shen N, et al. Extensive polymorphism observed in HIV - 1 clade B protease gene using high density oligonucleotides arrays. Nature Medicine, 1996, 2: 753

100. Sapolsky R , Lipshutz RJ. Mapping genomic library clines using oligonucleotide arrays. Genomics, 1996, 33: 445

101. Fodor SP, Read JL, Pirrung Mc, et al. Light-directed, spatially addressable parallel chemical synthesis. Science, 1991, 251(4995): 767

102. Sheppard CJR, Choudhury A. Image formation in the scanning microscope. J. Optia Acta, 1977, 24(10): 1051

103. Wodicka L, Dong H, Mittmann M, et al. Genome - wide expression monitoring in Saccharomyces cerevisiae. Nature Biotechnology, 1997, 15: 1359

104. John I Bell. Single nucleotide polymorphisms and disease gene mapping. Arthritis Res 2002, 4 (suppl 3): S273

105. Hoh J, Wille A, Ott J. Trimming weighting and grouping SNPs in human case - control association studies. Genome Research, 2001, 11(12) , 2115

106. Irizarry K, Kustanovich V, Li C , et al. Genome - wide analysis of single - nucleotide polymorphisms in human expressed sequences. Nature genetics, 2000, 26(2): 233

107. Ellis MC. "Spot - On" SNP Genotyping. Genome Res, 2000, 10 (7), 895

108. Schork NJ, Fallin D, Lanchbury S. Single nucleotide polymorphisms and the future of genetic epidemiology. Clin Genet, 2000, 58(4): 250

109. Ye S, Dhillon S, Ke X, et al. An efficient procedure for genotyping single nucleotide polymorphisms, Nucleic Acids Res, 2001, 29(17): E88

110. Hacia JG, Brody LC, Chee MS, et al. Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two - colour fluorescence analysis. Nat Genet , 1996, 14: 441

111. Chee MR, Yang E, Hubbell, et al. Accessing genetic information with high - density DNA arrays. Science, 1996, 274: 610

112. Head SR, Rogers YH, Parikh K, et al. Nested genetic bit analysis (N - GBA) for mutation detection in the p53 tumor suppressor gene. Nucleic Acids Res, 1997, 25: 5065

113. Sosnowski RG, Tu E, Butler WF, et al. Rapid determination of single base mismatch mutations in DNA hybrids by direct electric field control. Proc Natl Acad Sci USA, 1997, 94: 1119

114. Drmanac S, Kita D, Labat I, et al. Accurate sequencing by hybridization for DNA diagnostics and individual genomics. Nat Biotechnol, 1998, 16: 54

115. Hacia JG, Edgemon K, Sun B, et al. Two color hybridization analysis using high density oligonucleotide arrays and energy transfer dyes. Nucleic Acids Res , 1998, 26: 3865

116. Gerry NP, Witowski NE, Day J, et al. Universal DNA microarray method for multiplex detection of low abundance point mutations. J Mol Biol, 1999, 292: 251

117. Favis R, Barany F. Mutation detection in *K* - ras, *BRCA1*, *BRCA2*, and p53 using PCR/LDR and a universal DNA microarray. Ann N Y Acad Sci, 2000, 906: 39

118. Schena M, Shalon D, Heller R, et al. Parallel human genome analysis: microarray - based expression monitoring of

1 000 genes. Proc Natl Acad Sci USA, 1996,93(20):10614

119. Bubendorf L, Kolmer M, Kononen J, et al. Hormone therapy failure in human prostate cancer: analysis by complementary DNA and tissue microarrays. J Natl Cancer Inst, 1999, 91(20): 1758

120. Kitahara O, Furukawa Y, Tanaka T, et al. Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser - capture microdissection of tumor tissues and normal epithelia. Cancer Res, 2001, 61(9): 3544

121. Xu J, Stolk JA, Zhang X, et al. Identification of differentially expressed genes in human prostate cancer using subtraction and microarray. Cancer Res, 2000, 60(6): 1677

122. Villaret DB, Wang T, Dillon D, et al. Identification of genes overexpressed in head and neck squamous cell carcinoma using a combination of complementary DNA subtraction and microarray analysis. Laryngoscope, 2000, 110(3 Pt 1): 374

123. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999 Oct 15;286(5439):531

124. Sclaru FM, Zou T, Xu Y, et al. Global gene expression profiling in Barrett's esophagus and esophageal cancer: a comparative analysis using cDNA microarrays. Oncogene, 2002, 21(3): 475

125. Schuldiner O, Benvenisty N. A DNA microarray screen for genes involved in c - MYC and N - MYC oncogenesis in human tumors. Oncogene, 2001, 20(36): 4984

126. Hacia JG, Woski SA, Fidanza J, et al. Enhanced high density oligonucleotide array - based sequence analysis using modified nucleoside triphosphates. Nucleic Acids Res, 1998,26(21):4975

127. Solinas - Toldo S, Lampel S, Stilgenbauer S, et al. Matrix - based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer, 1997, 20: 399

128. Pinkel D, Segraves R, Sudar D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet, 1998, 20: 207

129. Zhou Y, Gwadry FG, Reinhold WC, et al. Transcriptional regulation of mitotic genes by camptothecin - induced DNA damage: microarray analysis of dose - and time - dependent effects. Cancer Res, 2002, 62(6):1688

130. van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature, 2002,415(6871):530

131. Kawai HF, Kaneko S, Honda M, et al. alpha - fetoprotein - producing hepatoma cell lines share common expression profiles of genes in various categories demonstrated by cDNA microarray analysis. Hepatology, 2001, 33:676

132. Huang Y, Uchiyama Y, Fujimura T, et al. A human hepatoma cell line expressing hepatitis C virus nonstructural proteins tightly regulated by tetracycline. Biochem Biophys Res Commun, 2001,281:732

133. Shirota Y, Kaneko S, Honda M, et al. Identification of differentially expressed genes in hepatocellular carcinoma with cDNA microarrays. Hepatology, 2001,33:832

134. Honda M, Kaneko S, Kawai H, et al. Differential gene expression between chronic hepatitis B and C hepatic lesion. Gastroenterology, 2001, 120: 955

135. Bigger CB, Brasky KM, Lanford RE. DNA microarray analysis of chimpanzee liver during acute resolving hepatitis C virus infection. J Virol, 2001, 75: 7059

136. Chen YH, Wang WC, Young KC, et al. Plastic microchip electrophoresis for PCR products of hepatitis C virus. Clinical Chemistry, 1999,45:1938

137. 刘伟,等.DNA芯片技术诊断乙、丙型病毒性肝炎的研究.东南大学学报(医学版),2002,21(1):75

138. Shackle NA, McGuinness PH, Abbott CA, et al. Identification of novel molecules and pathogenic pathways in primary biliary cirrhosis: cDNA array analysis of intrahepatic differential gene expression. Gut,2001,49: 565

139. Fodor SPA. Massively parallel genomics. Science. 1997, 277: 393

140. Southern E M, Maskos U, Elder J K. Analyzing and comparing nucleic acid sequences by hybridization to arrays of

olignucleotides: evaluation using experimental methods. Genomics, 1992, 13: 1008

141. Schena M. DNA microarrays: miniature assays. Seminar on techniques in DNA microarray. Beijing. 1999

142. Head SR, Parikh K, Rogers YH, et al. Solid - phase sequence scanning for drug resistance detection in tuberculosis. Mol Cell Probes, 1999, 13: 81

143. Niesters HG, Honkoop P, et al. Identification of more than one mutation in the hepatitis B virus polymerase gene arising during prolonged lamivudine treatment. J Infect Dis, 1998, 177: 1382

144. 马文丽,郑文岭.DNA芯片技术的方法与应用.广州:广东科技出版社,2002

145. Heller RA, Schena M, Chai A, et al. Discovery and analysis of inflammatory disease - related genes using cDNA microarrays. Proc Natl Acad Sci USA, 1997, 94(6): 2150

146. Anthony DC, Barbara EB. Identification of novel targets of immunosuppressive agents by cDNA - based microarray analysis. J Bio Chem, 2002, 227(6): 4465

147. Dove A. Haploinsufficiency drug screen. Nat Biotechnol, 1999, 17(4): 317

148. Cronin MT, Pho M, Dutta D. Utilization of new technologies in drug trials and discovery. Drug Metab Dispos, 2001, 29: 586

149. Gill P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. Int J Legal Med, 2001, 114(4 - 5): 204

150. Divne AM. A novel DNA microarray system for analysis of limited forensic evidence material. In: Proceedings from the 19th International Society Forensic Genetics, 2001, 30

151. Moretti TR, Baumstark AL, Defenbaugh DA, et al. Validation of STR typing by capillary electrophoresis. J Forensic Sci, 2001, 46(3): 661

152. Kallioniemi OP, Wagner U, Kononen J, et al. Tissue microarray technology for high - throughput molecular profiling of cancer. Hum Mol Genet, 2001, 10(7): 657

153. McGlennen RC. Miniaturization technologies for molecular diagnostics. Clin Chem, 2001, 47(3): 393

154. Borrebaeck CA. Antibodies in diagnostics - from immunoassays to protein chips. Immunol Today, 2000, 21: 379

155. Brody EN, Gold L. Aptamers as therapeutic and diagnostic agents. J Biotechnol, 2000, 74: 5

156. Coombes BK, Mahony JB. cDNA Array Analysis of Altered Gene Expression in Human Endothelial Cells in Response to Chlamydia pneumoniae Infection. Infection and Immunity, 2001, 69(3): 1420

157. Adam TW, George FS, Richard AM. High - speed DNA genotyping using microfabricated capillary elecrophoresis chips. Anal Chem, 1997, 69, 2181

158. Christel LA, Petersen K, McMillan W, Northrup MA. Rapid, automated nucleic acid probe assays using silicon microstructures for nucleic acid concentration. J Biomech Eng 1999 Feb; 121(1): 22

159. Hammond HA, Li Jin, Zhong Y. Evaluation of 13 short tandem repeat loci for use in personal identification application. Am J Hum Genet, 1994, 55: 175

160. Markx GH, Huang Y, Zhou XF, et al. Dielectrophoretic characterization and separation of micro - organisms, Microbiology, 1994, 140: 585

161. Rohde E, Vogt C, Heineman WR. The analysis of fountain pen inks by capillary electrophoresis with ultraviolet/visible absorbance and laser - induced fluorescence detection. Electrophoresis, 1998, 19(1): 31

162. Cheng J. Microchip - based Devices for Molecular Diagnosis of Genetic Diseases. Molecular Diagnosis, 1996, 1: 183

163. 李民,周宗灿.蛋白质芯片.生命的化学,2001,21(2):156

164. 王燕,刘昕,缪金明,等.细胞膜表面标志CD系列抗体蛋白芯片制备及应用研究.第三军医大学学报,2002,24(1):43

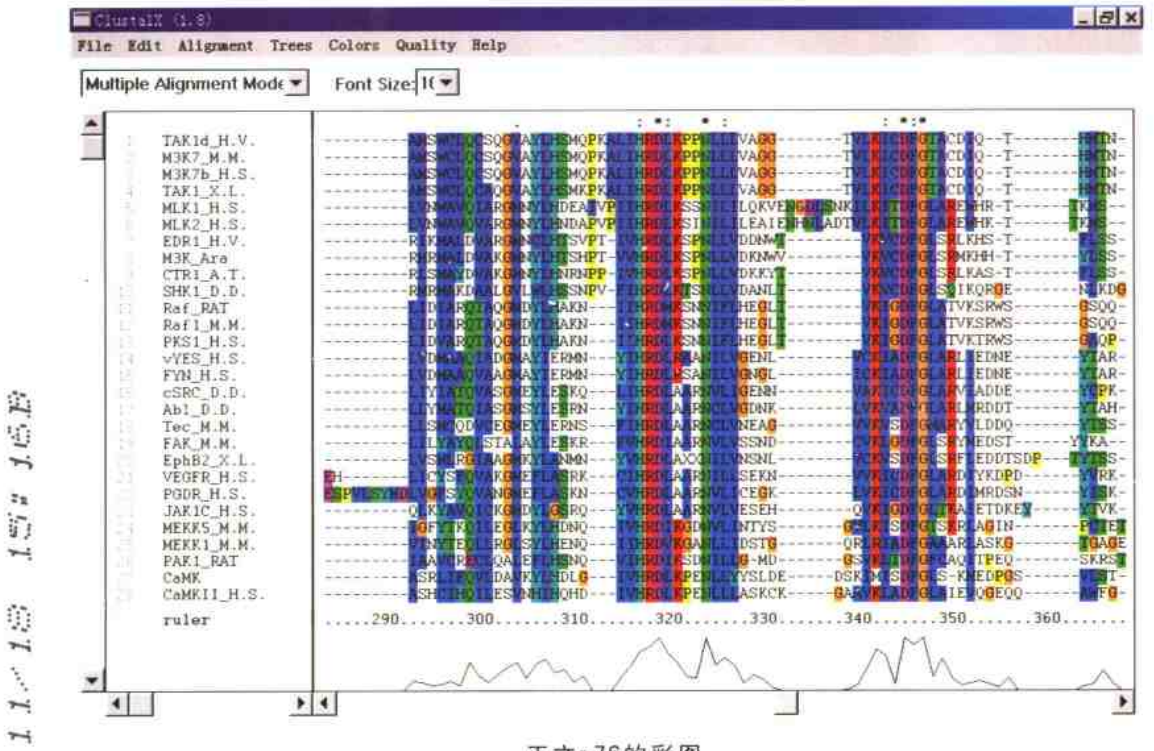
165. 赵芷音,程永樟,陈岳明.用芯片研究丙型肝炎病毒不同功能区抗体.中华医院感染学杂志,2001,11(5):327

166. Zhu H, Michael S. Protein array and microarrays. Current Opinion in Chemical Biology, 2001, 5: 40

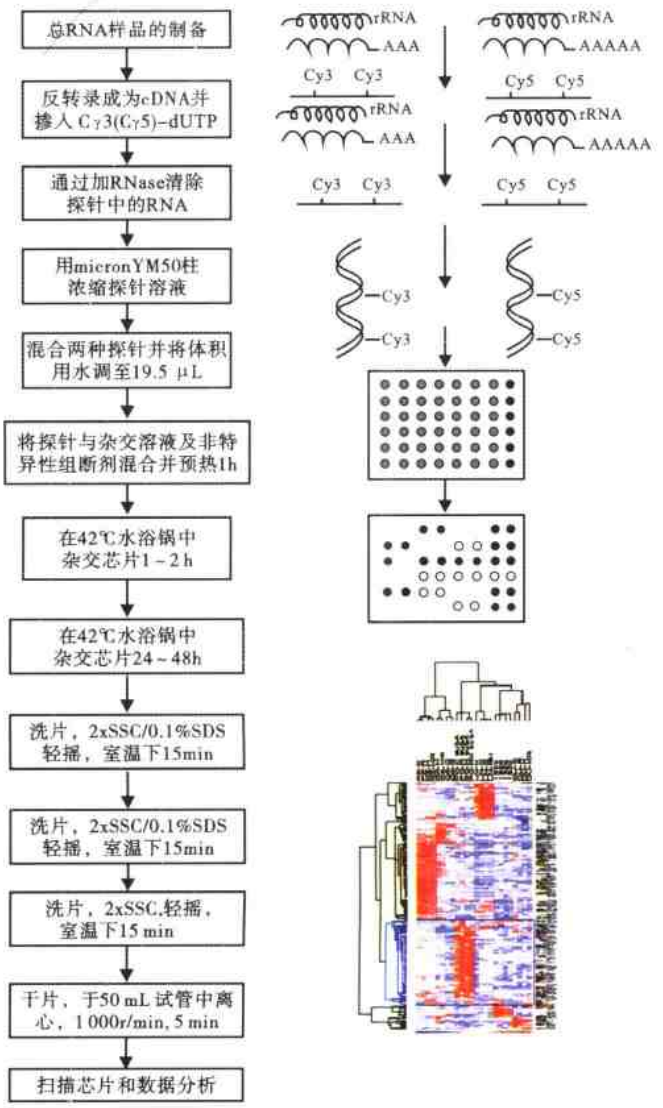
167. Dolores J, Cahill. Protein and antibody arrays and their medical applications. Journal of Immunological Methods, 2001,250:81

168. Eric TF, Vanitha T, Scott RW, et al. Protein biochips for differential profiling. Current Opinion in Biotechnology, 2001,12:65

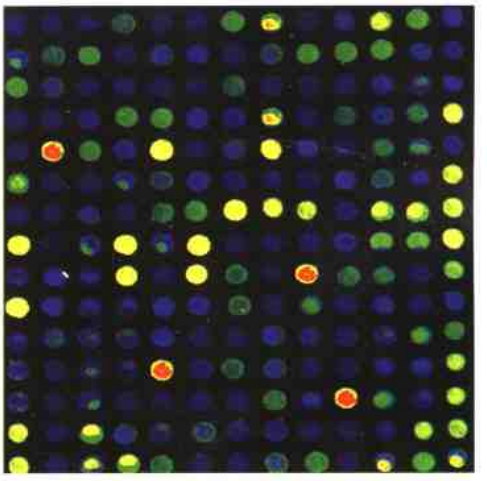
169. Wright GL, Cazaves LH, Leung SM, et al. Protein Chip Surface enhanced laser desorption/ionization(SELDI) mass spectrometry: a novel protein biochip technology for detection of biomarker in complex protein mixtures. Prostate Cancer and Prostatic Disease, 2000,2:264



正文p76的彩图



正文p218图15-13的彩图



正文p236图15-17的彩图